Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays on Econometric Methods for Duration Data Analysis

Autor:

Andrés F. García Suaza

Director/es:

Miguel A. Delgado

DEPARTAMENTO DE ECONOMÍA

Getafe, Junio 2016

To my parents, my bother,

and my unconditional wife Nathalia

Acknowledgements

First, I would like to express my gratitud to my advisor, Miguel A. Delgado for his inmense contribution to my academic carrer. During these years I learnt no only the basis on the academic work, but also the importance and the value of the persistence and the dedication. I also want to thank to all the professors I had during my stay at Carlos III, they have been fundamental in my training. In special to professor Winfried Stute, Carlos Velasco, Juan Carlos Escanciano.

I also want to mention my colleagues, Pedro Sant'Anna, Mehdi Hamidisahneh, Christos Mavridis, Meng Lee, Dionisis Karavidas, Pascualina Arca, Gonzalo Zunino, Sebastian Panthoefer, Federico Masera, Jose Carbo, Federico Curci and Hasin Yousaf, share this experience with you guys has been great!. And of course to my classmate, flatmate, officemate and friend Salvatore Lo Bello, thanks for your support and nice experiencies during these years.

Although, this dissertation is result of the last 5 years, I want to mention those many professors and mentors that always supported and encouraged me, Sergio Restrepo, Henry Laniado, Wilman Gomez, Mauricio Alvear, Pedro Ramirez, Luis Eduardo Arango, Carlos Esteban Posada, Jose Eduardo Gomez, Luis Melo, Juan Carlos Guataqui, Luis Fernando Gamboa, Ricardo Arguello, Jesus Otero, Hernando Zuleta and Fernando Jaramillo.

There are many people that made easier my life in the university. I want to thank the Master and PhD directors Angel Hernando-Veciana, Natalia Fabra, Andres Erosa and Luisa Fuster, and all administrative staff specially to Arancha Alonso y Angelica Aparicio. This project could not be done without my follow supports Banco de la República, Colfuturo and Universidad del Rosario.

Finally, I would like to thank my family. My parents and my brother and aunt

Angelica, they were my motivation and support no matter the distance. The last mention for my "costillita", Nathalia Florez. Thanks for your unconditional love, patience and for let me be part of the same team.

Abstract

In economic analysis is usual to find that the outcome of interest represents the duration until an event occurs, e.g. the duration until getting a job, the firms' lifetime, among others. The major challenge to analyze duration or survival data is the presence of censoring. The most of the existing survival models usually assume a parametric or semiparametric conditional hazard function. This thesis is formed by three chapters regarding alternative semiparametric estimation methods suitable for survival times observed under random censoring that do not require assumptions on the underlying duration distribution. These methods are motivated and applied in the context of unemployment duration studies.

Chapter 1 studies counterfactual decomposition methods. Existing inference procedures applicable when data is fully observed, might produce missleading conclussions. This may explain the lack of decomposition exercises for variables related to duration outcomes, typically observed under right censoring. We propose two decomposition methods that consider the presence of this kind of censoring. First, under suitable restrictions on the censoring mechanism, we provide an Oaxaca-Blinder type decomposition method of the mean in a nonparametric context. Consistent estimation of the decomposition components is based on a prior estimator of the joint distribution of duration and covariates. Secondly, we consider a method that makes possible to decompose other distributional features, such as the median or the Gini coefficient. To do so, weaker assumptions on the censoring nature are needed, but it is required to introduce restrictions on the functional form of the conditional distribution of duration given covariates. We provide formal justification for asymptotic inference and study the finite sample performance through Monte Carlo experiments. Finally, we apply the proposed methodology to the analysis of unemployment duration gaps in Spain. This study

suggests that factors beyond the workers' socioeconomic characteristics play a relevant role in explaining the difference between several unemployment duration distribution features such as the mean, the probability of being long term unemployed and the Gini coefficient.

Chapter 2 proposes inference procedures on distributional regression models in the context of survival analysis. These models generalize classical survival models to a situation where slope coefficients depend on duration time. We formally justify asymptotic inferences on the varying coefficients under weak regularity conditions, similar to those needed when data is not censored. Finite sample properties of the proposed inference procedures are studied by means of Monte Carlo experiments. Finally, proposed method is implemented in two empirical exercises using US data. First, we study the effect of unemployment benefits on unemployment duration; and secondly we perform a counterfactual decomposition in the context of the recent Great Recession using US data.

Chapter 3 adapts the generalized method of moments (GMM) to estimating parameters identified by moment restrictions involving survival time observed under right random censoring. When the underlying nonparametric joint distribution of survival time and the rest varibles can be identified under random censoring, the moment restrictions can be consistently estimated by weighting averages, which form a basis for the proposed GMM. Under classical assumptions in GMM estimation, we show consistency and asymptotic normality, and provide the optimal weighting matrix that maximizes relative efficiency. Finite sample properties are studied using a Monte Carlo expertiment of a linear in parameter structural model.

Contents

List of Tables

List of Figures

1 Oaxaca-Blinder type Counterfactual Decomposition Methods for Duration Outcomes

1.1 Introduction

From Oaxaca (1973) and Blinder (1973) contributions (OB henceforth), counterfactual decomposition technique became a popular research tool for economic analysis. This consists in decomposing the difference between the means of two subpopulations into counterfactual components based on observed characteristics. On some occasions, the decomposition of the mean may not be enough for studying the difference in the outcome of interest, so that the decomposition of other distributional features, such as the median or Gini coefficient, has also been proven useful. In this context, Freeman (1980); Juhn et. al. (1991); DiNardo et. al. (1996); Machado and Mata (2005) and Chernozhukov et. al. (2013) developed further decomposition techniques going beyond the mean. See Fortin et. al. (2011) for a comprehensive review.

The aforementioned procedures, designed for the case of fully observed data, have been widely used in the analysis of microdata, mainly in labor economics. For instance, there is a large literature devoted to studying the gender wage gap (see Oaxaca, 1973; Blinder, 1973; Cain, 1986; Oaxaca and Ransom, 1999; Machado and Mata, 2005; O'Neill and O'Neill, 2006; Blau and Kahn, 1992) or the increase in the US wage dispersion in the 1980's (see Juhn et. al., 1991, 1993; DiNardo et. al., 1996; Melly, 2005). However, other relevant outcomes, such as unemployment duration, have not received so much attention, possibly due to the absence of decomposition methods for censored variables.

Collecting duration data requires following individuals over time. In this context, censoring might occur because individuals either do not change their status during the follow-up period or withdraw before the end of the study. For instance, in the case of unemployment duration, it is not possible to observe the complete unemployment duration for those individuals that are still unemployed at the last follow-up period or for those who leave the labor force.

Existing literature concerning the decomposition of the unemployment duration has focused on the study of the average conditional hazard rate. However, unless a linear specification is used, the resulting decomposition components do not correspond to the observed total difference, and consequently, their interpretation cannot be done in the line of the OB decomposition[1]. Therefore, we consider decomposition methods suitable for several parameters related to a duration outcome. These methods allow to deal with covariates of diverse nature (continuous and discrete) and results can be interpreted as in the no censoring case, and also are simple to implement using conventional software[2].

In particular, we provide a regression-based method, analogous to the classical OB for the decomposition of the mean difference, in a nonparametric context. Consistent estimation of the means in the two subpopulations, and the corresponding decomposition components is based on a prior estimation of the joint distribution of duration outcome and covariates. We introduce a multivariate version of the joint cumulative hazard function which allows to obtain nonparametric identification of the joint distribution of duration and covariates, under classical assumptions in survival analysis literature. This approach allows to keep computational tractability and avoids the curse of dimensionality and the need of choosing tuning parameters. Under standard regularity conditions, asymptotic results for making

[1]For outcomes typically analyzed through non-linear models (e.g. dummy variables, count data and duration), OB decomposition methods based on linear approximations have been proposed (see Bauer and Sinning, 2008, and Powers and Yun, 2009). Our approach does not rely in any linear approximation.

[2]All codes used are available upon request.

statistical inference on the decomposition components are provided. However, the use of bootstrapping techniques can also be implemented.

Additionally, we consider an alternative method to perform counterfactual decomposition for parameters different to the mean. This encompasses real-valued parameters such as the median or well known inequality measures like the Gini coefficient, but also function-valued parameters such as the hazard function, the distribution function or the Lorenz Curve. This method is built on weaker restrictions on the censoring mechanism, but some knowledge on the functional form of the underlying conditional distribution of duration given covariates is required. The crucial issue to achieve the decomposition components is the estimation of the unconditional distribution of potential outcomes. We discuss the estimation of unconditional distribution of the counterfactual outcomes through the specification of conditional hazard rates, considering the proportional hazard models. But other specifications, e.g., a quantile regression or a fully parametric model, are also possible. Consistency and inference validity results follows the arguments of Chernozhukov et. al. (2013) for the estimation of counterfactual distributions.

The proposed methodologies are used to study unemployment duration gender gaps in Spain during the period 2004-2007 using data from the European Survey on Income and Living Conditions (SILC). We investigate the unemployment duration gender gap in two different scenarios: the duration until leaving unemployment and the duration until getting a job. To do so, we perform counterfactual decomposition for several parameters of the unemployment duration to quantify to what extent the gender gap is explained by observable socioeconomic factors, like individual and household background. Particularly, we study the mean unemployment duration, the probability of being long term unemployed and the Gini coefficient. Our findings reveal that differences in observable characteristics play a minor role to explain gender gaps in unemployment duration, but it is statistically significant to explain the duration until getting a job.

Our contribution is twofold. First, from the methodological point of view, we extend classical methods to perform decomposition to the case of right censored outcomes. We discuss useful issues for practitioners such as the implications of the identification assumptions, estimation and inference. Based on Monte Carlo experiments, we also show the consequence of neglecting the censoring and finite sample properties of the proposed methods. Although we analyze unemployment data, our methods enable to analyze other relevant economic outcomes denoting duration as firm's lifetime, bank failure, exporting relation spells, school dropout, among others.

Secondly, our empirical results contribute to deepen the understanding on determinants of unemployment gender differentials. There is a large literature that explore the presence of gender gaps in unemployment rate (see c.f. Niemi, 1974; Johnson, 1983; Azmat et. al., 2006, and Queneau and Sen, 2007), or in the unemployment exit rate (see c.f. Eusamio, 2004; Ortega, 2008, and Tansel and Tasci, 2010). We consider in our analysis other dimensions of the gender gaps such as severity and inequality in terms of unemployment.

The rest of the chapter is organized as follows. The next section introduces the implementation of the OB method for the censored data case in a full nonparametric context, and provides sufficient conditions to perform valid inferences on the counterfactual decomposition components. Third section discusses the decomposition of other distributional features under a semiparametric specification. Fourth section studies the finite sample properties through Monte Carlo simulations. Fifth section applies the proposed methodologies to analyze unemployment duration gender gaps in Spain. The last section is devoted to final remarks. Some mathematical details and further discussion on technical results are presented in the Appendix.

1.2 Nonparametric Oaxaca-Blinder Decomposition under Censoring

Consider a $\mathbb{R}^+ \times \mathbb{R}^k \times \{0,1\} - valued$ random vector (T, X, D) related to the population under study, where T denotes duration outcome, X a $k \times 1$ vector of characteristics (including a constant) and D a dummy variable identifying two subpopulations. For instance, T may be unemployment duration, X relevant socioeconomic characteristics and D a dummy variable for gender.

The difference between the means of the two subpopulations, denoted by $\Delta_T^\mu = \mu_T^{(1)} - \mu_T^{(0)}$, with $\mu_T^{(\ell)} = \mathbb{E}(T \mid D = \ell)$, $\ell = \{0,1\}$, can be expressed in terms of the best linear predictors for each subpopulation ℓ. That is:

$$\Delta_T^\mu = \beta_1' \mu_X^{(1)} - \beta_0' \mu_X^{(0)} \tag{1.1}$$

where, $\mu_X^{(\ell)} = \mathbb{E}(X \mid D = \ell)$,

$$\beta_\ell = \underset{b \in \mathbb{R}^k}{\arg\min} \quad \mathbb{E}\left[(T - b'X)^2 \mid D = \ell\right] = \mathbb{E}(XX' \mid D = \ell)^{-1} \mathbb{E}(XT \mid D = \ell)$$

and $\mathbb{E}$ is the expectation operator and A' denotes the transpose of A.

Oaxaca (1973) and Blinder (1973) exploit this fact to rearrange Equation (1.1) as

$$\Delta_T^\mu = (\beta_1 - \beta_0)' \mu_X^{(1)} + \beta_0' \left(\mu_X^{(1)} - \mu_X^{(0)}\right) = \Delta_S^\mu + \Delta_C^\mu. \tag{1.2}$$

This is the classical Oaxaca-Blinder decomposition (OB decomposition, henceforth), where the term Δ_S^μ , known as *the structure effect*, is interpreted as the difference explained by the effect (return) of the explanatory variables on T, while

Δ_C^μ, known as *the composition effect*, is the part of the mean difference explained by the difference in the characteristics.

The crucial ingredient in this counterfactual decomposition is $\beta_0'\mu_X^{(1)}$, i.e. the best predictor of T in subpopulation 0 given $X = \mathbb{E}(X \mid D = 1)$. Intuitively, this is the average of the counterfactual outcome denoted by $T^{(0,1)}$, which represents the potential outcome if individuals in subpopulation 0 were endowed with characteristics of subpopulation 1. Or the same, the potential outcome in subpopulation 1 if they would face the circumstances or schedule of the other subpopulation[3]. Indeed, the label *counterfactual* comes from the fact that this outcome cannot be directly observed in the data[4].

In order to identify $\mu_T^{(0,1)}$, the average of $T^{(0,1)}$, it is necessary to impose some restrictions. Assumption 1.1 below summarizes the identification conditions usually considered in the decomposition methods literature (see Fortin et. al., 2011 for further discussion).

Assumption 1.1 *Let $\varepsilon^{(\ell)}$ be the best linear predictor error for subpopulation ℓ, i.e. $\varepsilon^{(\ell)} = \left(T^{(\ell)} - \beta_\ell' X^{(\ell)}\right)$, with $T^{(\ell)}$ and $X^{(\ell)}$ the outcome variable and covariates of the corresponding subpopulation. The following conditions hold:*

a. Overlapping support: if $\mathcal{X} \times \mathcal{E}$ denotes the support of observables and unobservable characteristics of the underlying population, then $\left(X^{(0)}, \varepsilon^{(0)}\right) \cup \left(X^{(1)}, \varepsilon^{(1)}\right) \in \mathcal{X} \times \mathcal{E}$.

b. The only possible counterfactual outcome for an individual that belongs to sub-

[3] There are alternative ways to rearrange Equation (1.2) to get these decomposition components. We focus in this representation in sake of simplicity, but analogous results are obtained for the other representations of the OB decomposition.

[4] Counterfactual analysis, as a concept, has been used in a very philosophical way in many sciences (Lewis, 1973). In social science, and particularly in economics since the seminal contribution by Rubin (1974), counterfactual analysis has served to establish a natural framework for studying causal relations (For discussion on the use of counterfactuals in quantitative analysis see Dawid, 2000; Cartwright and Reiss, 2004; Höfler, 2005, and Pearl, 2009).

population i is $T^{(j)}$, with $i,j \in \{0,1\}$ and $i \neq j$.

c. Conditional independence of the group label and unobservables: $D \perp \varepsilon | X$.

Assumption 1.1.a is the classical common support condition. Assumption 1.1.b, known as *simple counterfactual treatment*, rules out the existence of other potential outcomes besides $T^{(0)}$ and $T^{(1)}$. Lastly, Assumption 1.1.c is the classical ignorability (*unconfoundedness*) condition, which ensures that the distribution of unobservables is the same across subpopulations[5]. Conditions 1.1.b-1.1.c imply that the conditional distributions of the outcome and the unobservables given covariates remain unaltered when the distribution of covariates varies. This invariance property is the key assumption to identify the mean of the counterfactual outcomes by combining parameters from the two subpopulations, in particular, $\mu_T^{(0,1)} = \beta_0' \mu_X^{(1)}$.

Given a random sample $\{T_i, X_i, D_i\}_{i=1}^n$ from (T, X, D), the OB decomposition is estimated by:

$$\bar{\Delta}_T^\mu = \bar{\Delta}_S^\mu + \bar{\Delta}_C^\mu = \left(\bar{\beta}_1 - \bar{\beta}_0\right)' \bar{\mu}_X^{(1)} + \bar{\beta}_0' \left(\bar{\mu}_X^{(1)} - \bar{\mu}_X^{(0)}\right),$$

where $\bar{\mu}_X^{(\ell)} = n_\ell^{-1} \sum_{i=1}^n X_i 1_{\{D_i = \ell\}}$, and

$$\bar{\beta}_\ell = \underset{b \in \mathbb{R}^k}{\arg\min} \sum_{i=1}^n \left(T - b'X\right)^2 1_{\{D_i = \ell\}}$$

with $n_\ell = \sum_{i=1}^n 1_{\{D_i = \ell\}}$, $\ell = \{0,1\}$ and $1_{\{A\}}$ denoting the indicator function of the event A.

However, in practice, these estimators are infeasible when T is observed under censoring. In the context of duration data, censoring appears due to lack of follow-up. If individuals are observed along a fixed period, complete durations are not

[5]Even though identification of decomposition factors is given by analogous assumptions to those used in the policy evaluation literature, the causal interpretation requires stringent conditions on the nature of the treatment D and the control variables X.

always available because either the relevant event did not occur at the end of the observation period, or the individual abandoned the study. For instance, in an unemployment duration study some individuals might be still unemployed at the end of the follow-up period, while others leave the labor force. Under these circumstances, the observed sample is characterized by $\{Y_i, X_i, \delta_i, D_i\}_{i=1}^n$, n copies of the random vector (Y, X, δ, D), where $Y = \min(T, C)$, $\delta = 1_{\{T \leq C\}}$ and C denotes the censoring times. In this case, the estimator $\bar{\Delta}_T^\mu$, based on observed durations Y, turns out biased.

To construct a corrected version of the OB decomposition with presence of censoring, consider the joint distribution of (T, X) given that $D = \ell$, $F^{(\ell)}(t, x) = \mathbb{P}(T \leq t,\ X \leq x | D = \ell)$, where henceforth $\leq$ is coordinatewise. Notice that, for $\ell = \{0, 1\}$, we can express $\mu_T^{(\ell)} = \int_{\mathbb{R}} t dF^{(\ell)}(t, \infty)$, $\mu_X^{(\ell)} = \int_{\mathbb{R}^k} x dF^{(\ell)}(\infty, x)$, and

$$\beta_\ell = \underset{b \in \mathbb{R}^k}{\arg\min} \int (t - b'x)^2 \, dF^{(\ell)}(t, x).$$

In fact, in the absence of censoring, $\bar{\Delta}_T^\mu$ is the sample analog of Δ_T^μ, where F is replaced by the sample version

$$\bar{F}^{(\ell)}(t, x) = \frac{1}{n_\ell} \sum_{i=1}^n 1_{\{T_i \leq t, X_i \leq x, D_i = \ell\}}.$$

Under censoring, a consistent estimator of $F^{(\ell)}$ can be obtained by exploiting its representation in terms of the cumulative, or integrated, hazard function. Consider, in the context of an unemployment study, the probability that an individual, taken at random at time t from the subpopulation ℓ that belongs to the group of individuals with characteristics $\{X \in B\}$, finds a job before $t+h$. This probability can be written as,

$$\begin{aligned}
\mathbb{P}\left(t \leq T < t+h, X \in B \mid T \geq t, D=\ell\right) &= \frac{\mathbb{P}\left(t \leq T < t+h, X \in B, D=\ell\right)}{\mathbb{P}\left(T \geq t, D=\ell\right)} \\
&= \int_{\{X \in B\}} \frac{F^{(\ell)}\left(t+h, dx\right) - F^{(\ell)}\left(t-, dx\right)}{1-F^{(\ell)}(t-, \infty)}
\end{aligned}$$

where for any generic function J, $J\left(t-\right)=\lim_{x\uparrow t} J(x)$, and $B \in \boldsymbol{\beta}^k$ the smallest sigma algebra in $\mathbb{R}^k$. For practical purposes, take $B=(-\infty, x]$.

Suppose that there exists a function λ such that

$$\frac{F^{(\ell)}\left(t+h, x\right)-F^{(\ell)}\left(t-, x\right)}{1-F^{(\ell)}(t-, \infty)} = h\lambda^{(\ell)}\left(t, x\right) \text{ as } h \rightarrow 0.$$

This function $\lambda^{(\ell)}\left(., x\right)$ is the hazard function for individuals of subpopulation ℓ with $\{X \leq x\}$. The function $\lambda^{(\ell)}\left(t, x\right)$ can be interpreted as the instantaneous probability that an individual belonging to subpopulation ℓ with characteristics $\{X \leq x\}$, leaves from unemployment immediately after moment t.

Therefore, the associated cumulative hazard can be defined as

$$\Lambda^{(\ell)}\left(t, x\right)=\int_0^t \frac{F^{(\ell)}\left(d\bar{t}, x\right)}{1-F^{(\ell)}(\bar{t}-, \infty)}, \tag{1.3}$$

and if $\lambda^{(\ell)}$ exists, $\Lambda^{(\ell)}\left(t, x\right)=\int_0^y \lambda^{(\ell)}\left(\bar{t}, x\right) d\bar{t}$. Using the fact that any distribution function can be expressed in terms of the corresponding integrated hazard (see Gill, 1980; Gill and Johansen, 1990, and Shorack and Wellner, 2009, p. 301 Proposition 1, for details), we have

$$1-F^{(\ell)}(t, x)=\exp\left\{-\Lambda^{(\ell)c}(t, x)\right\} \prod_{\bar{t} \leq t}\left[1-\Lambda^{(\ell)}\left(\{\bar{t}\}, x\right)\right] \tag{1.4}$$

where $\Lambda^{(\ell)c}$ is the continuous part of $\Lambda^{(\ell)}$, and for any generic function J, $J\{t\}=J(t)-J(t-)$. Therefore, $F^{(\ell)}(t, x)$ can be estimated by plugging-in a proper estimator of $\Lambda^{(\ell)}$. Because of the presence of censoring, the identification of $\Lambda^{(\ell)}$

requires imposing restrictions on the censoring mechanism. To do so, we consider Assumption 1.2.

Assumption 1.2 *The following conditions hold:*
a. $\mathbb{P}(T \leq t, C \leq c \mid D = \ell) = \mathbb{P}(T \leq t \mid D = \ell)\, \mathbb{P}(C \leq c \mid D = \ell)$.
b. $\mathbb{P}(T \leq C \mid T, X, D) = \mathbb{P}(T \leq C \mid T, D)$.

This assumption has been widely used in survival analysis (c.f. Stute, 1993, 1996, 1999; Uña-Álvarez and Rodríguez-Campos, 2004; Sanchez-Sellero et. al., 2005 and Sant'Anna, 2016). Assumption 1.2.a. is the classical independence assumption that guarantees identification of the marginal distribution of survival times (c.f. Peterson, 1977). In turn, Assumption 1.2.b. states the relation between the censoring mechanism and the covariates so that, given the actual survival times T, the covariates do not provide any further information on whether censoring occurs (see Stute, 1993 for further discussion). In this framework, potential dependence between C and X is allowed, and of course, it is also held when C is independent of (T, X).

Then, under Assumption 1.2 we can express $\Lambda^{(\ell)}$ in terms of the following sub-distributions:

$$\begin{aligned} H^{(\ell)}(t) &= \mathbb{P}(Y \leq t | D = \ell), \text{ and} \\ H_{11}^{(\ell)}(t, x) &= \mathbb{P}(Y \leq t, X \leq x, \delta = 1 | D = \ell). \end{aligned}$$

Proposition 1.1 *Under Assumption 1.2, the joint cumulative hazard function can be written as:*

$$\Lambda^{(\ell)}(t, x) = \int_0^t \frac{H_{11}^{(\ell)}(d\bar{t}, x)}{1 - H^{(\ell)}(\bar{t}-)}.$$

The sample analogs of $H^{(\ell)}(t,\ell)$ and $H_{11}^{(\ell)}(t,x)$ are given by

$$\hat{H}^{(\ell)}(t) = n_\ell^{-1}\sum_{i=1}^{n} 1_{\{Y_i \le t, D_i=\ell\}} \quad \text{and} \quad \hat{H}_{11}^{(\ell)}(t,x) = n_\ell^{-1}\sum_{i=1}^{n} 1_{\{Y_i \le t, X_i \le x,\ D_i=\ell, \delta_i=1\}}$$

and hence, $\Lambda^{(\ell)}(t,x)$ is estimated by

$$\hat{\Lambda}^{(\ell)}(t,x) = \int_0^t \frac{\hat{H}_{11}^{(\ell)}(d\bar{t},x)}{1-\hat{H}^{(\ell)}(\bar{t}-)} = \sum_{i=1}^{n_\ell} \frac{1_{\{Y_i\le t, X_i \le x, D_i=\ell, \delta_i=1\}}}{n_\ell - R_i^{(\ell)} + 1}$$

where $R_i^{(\ell)} = n_\ell \hat{H}^{(\ell)}(Y_i)$ is the rank of Y_i provided that $i-th$ individual belongs to subpopulation ℓ.

As a consequence, the joint distribution can be estimated by

$$\begin{aligned} \hat{F}^{(\ell)}(t,x) &= 1 - \prod_{\bar{t}\le t}\left[1-\hat{\Lambda}^{(\ell)}\left(\{\bar{t}\},x\right)\right] \qquad (1.5) \\ &= 1 - \prod_{Y_{i:n_\ell}^{(\ell)} \le t, X_{[i:n_\ell]}^{(\ell)} \le x} \left[1 - \frac{\delta_{[i:n_\ell]}^{(\ell)}}{n_\ell - R_i^{(\ell)} + 1}\right] \end{aligned}$$

where $Y_{1:n_\ell}^{(\ell)} \le Y_{2:n_\ell}^{(\ell)} \le \ldots \le Y_{n_\ell:n_\ell}^{(\ell)}$ are the order statistics of Y in subpopulation ℓ, i.e. $Y_{i:n_\ell}^{(\ell)} = Y_j$ if $R_j^{(\ell)} = i$. In case of ties, it is considered as if uncensored observations precede the censored observations, and other kind of ties are ordered arbitrarily. And for any $\{\xi_i\}_{i=1}^{n_\ell}$, $\xi_{[i:n_\ell]}^{(\ell)}$ is the $i-th$ $\xi^{(\ell)}$-concomitant of $Y_{i:n_\ell}^{(\ell)}$, that is, $\xi_{[i:n_\ell]}^{(\ell)} = \xi_j$ if $Y_{i:n_\ell}^{(\ell)} = Y_j$. Equation (1.5) is the version of the Kaplan-Meier estimator (Kaplan and Meier, 1958) taking into account covariates. In fact, in absence of covariates, Equation (1.5) reduces to the classical product-limit version of the Kaplan-Meier estimator. Corollary 1.1 states an alternative representation of the estimator above.

Corollary 1.1 *Under Assumption 1.2, the estimator of the joint distribution 1.5* *can be written as,*

$$\hat{F}^{(\ell)}(t,x) = \sum_{i=1}^{n} W_i^{(\ell)} 1_{\left\{Y_{i:n_\ell}^{(\ell)} \leq t, X_{[i:n_\ell]}^{(\ell)} \leq x\right\}}$$

where

$$W_i^{(\ell)} = \frac{\delta_{[i:n_\ell]}^{(\ell)}}{n_\ell - R_i^{(\ell)} + 1} \prod_{j=1}^{i-1} \left[1 - \frac{\delta_{[j:n_\ell]}^{(\ell)}}{n_\ell - R_j^{(\ell)} + 1}\right].$$

The weights $\left\{W_i^{(\ell)}\right\}$ represent the mass attached to the $i-th$ order statistic $Y_{i:n_\ell}^{(\ell)}$. Intuitively, the role of these weights is to redistribute the mass across order statistics. That is, it is assigned a mass of $\frac{1}{n}$ to each observation. The status of each order statistic, $\delta_{[i:n_\ell]}^{(\ell)}$, is inspected until finds the first censored observation, whose attached mass is re-assigned uniformly to the remaining individuals to the right. And this process is repeated with the subsequent censored observations[6]. Note that $\hat{F}^{(\ell)}(t,x)$ assigns zero weight to censored observations and in absence of censoring, i.e. when $\delta_{[i:n_\ell]}^{(\ell)} = 1\ \forall i$, it reduces to the multivariate empirical distribution with $W_i^{(\ell)} = n_\ell^{-1}$.

Asymptotic properties of $\hat{F}^{(\ell)}(t,x)$ and the associated empirical integrals (known as *Kaplan Meier integrals*) of the form $\int \varphi(t,x)\, d\hat{F}^{(\ell)}(t,x)$, with φ an integrable function, have been studied by Stute (1993, 1996). This nonparametric estimator has advantages on other alternative estimators of the joint distribution. In particular, it does not rely on any shape restriction, is simple to compute, avoids the curse of dimensionality and the use of smoothers.

In this way, the OB decomposition under censoring (we call *Censored Oaxaca-Blinder*, COB hereafter) can be computed replacing $F^{(\ell)}$ by its sample analog $\hat{F}^{(\ell)}$.

[6] This estimator of the joint distribution can be interpreted as a Inverse-Probability-Weighting estimator in the lines of Horvitz and Thompson (1952). In fact, the weights can be obtained multiplying the status $\delta_{[i:n_\ell]}^{(\ell)}$ by the inverse of the probability of observing a failure (see Efron, 1967; Robins and Rotnitzky, 1992; Satten and Datta, 2001, for further discussion in the univariate case).

In particular, the total difference Δ_T^μ is estimated by:

$$\hat{\Delta}_T^\mu = \hat{\mu}_T^{(1)} - \hat{\mu}_T^{(0)}$$

where $\hat{\mu}_T^{(\ell)} = \sum_{i=1}^{n_\ell} W_i^{(\ell)} Y_{i:n_\ell}^{(\ell)}$, and the counterfactual decomposition components are

$$\hat{\Delta}_T^\mu = \left(\hat{\beta}_1 - \hat{\beta}_0\right)^T \hat{\mu}_X^{(1)} + \hat{\beta}_0^T \left(\hat{\mu}_X^{(1)} - \hat{\mu}_X^{(0)}\right) \tag{1.6}$$

where $\hat{\mu}_X^{(\ell)} = \sum_{i=1}^{n_\ell} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)}$, and $\hat{\beta}_\ell$ is estimated by the weighted least squares procedure[7] given by

$$\hat{\beta}_\ell = \underset{b \in \mathbb{R}^k}{\arg\min} \int (t - b'x)^2 d\hat{F}^{(\ell)}(t, x) = \underset{b \in \mathbb{R}^k}{\arg\min} \sum_{i=1}^{n_\ell} W_i^{(\ell)} (Y_{i:n_\ell}^{(\ell)} - b' X_{[i:n_\ell]}^{(\ell)})^2$$

As a general feature in the context of censored outcomes, consistency of the estimator in Equation (1.6) requires additional restrictions on the support of the duration outcome and censoring times (for details, see Stute and Wang, 1993; Stute, 1995; Sanchez-Sellero et. al., 2005). Define $F_T^{(\ell)}(t) = \mathbb{P}(T \leq t | D = \ell)$, i.e. $F_T^{(\ell)}(t) = F^{(\ell)}(t, \infty)$, and the distribution function of the censoring times as $G^{(\ell)}(t) = \mathbb{P}(C \leq t | D = \ell)$. Additionally, for a generic distribution $J^{(\ell)}(t)$ define the least upper bound as $\tau_J^{(\ell)} = \inf\left\{t : J^{(\ell)}(t) = 1\right\} \leq \infty$.

Assumption 1.3 *For $\ell = \{0, 1\}$, it holds that $\tau_{F_T}^{(\ell)} \leq \tau_G^{(\ell)}$.*

If $\tau_H^{(\ell)} = \tau_{F_T}^{(\ell)} \leq \tau_G^{(\ell)}$, estimators above are consistent over all the support. But, in the case when $\tau_H^{(\ell)} = \tau_G^{(\ell)} < \tau_{F_T}^{(\ell)}$, the inference is restricted to $\left(0, \tilde{Y}^{(\ell)}\right]$, $\tilde{Y}^{(\ell)}$

[7] There are other alternative regression methods for censored data (see Miller, 1976; Buckley and James , 1979; Koul et. al., 1981; Miller and Halpern, 1982; Ritov, 1990; Heuchenne and Keilegom, 2007); this approach provides a parsimonious method. For instance, it is flexible to compute functions involving both the duration outcome and the covariates useful for making statistical inference, and β_ℓ is simpler to compute, avoiding the use of iterative methods and smoothers.

$\leq Y^{(\ell)}_{n_\ell:n_\ell} < \tau^{(\ell)}_H$, otherwise the estimates are typically downward biased[8] (c.f. Gill, 1980; Mauro, 1985; Stute, 1994).

Lastly, Proposition 1.2 and Corollary 1.2 provide the basis to perform statistical inference on the counterfactual decomposition components in Equation (1.6). First, define:

$$\gamma_0^{(\ell)}(t) = \exp\left\{\int_0^{t-} \frac{H_0^{(\ell)}(d\bar{t})}{1 - H^{(\ell)}(\bar{t})}\right\},$$

$$\gamma_1^{(\ell)}\left(t; \varphi^{(\ell)}\right) = \frac{1}{1 - H^{(\ell)}(t)} \int 1_{\{t<\bar{t}\}} \varphi^{(\ell)}(\bar{t}, x) \gamma_0^{(\ell)}(\bar{t})\, dH_{11}^{(\ell)}(\bar{t}, x),$$

$$\gamma_2^{(\ell)}\left(t; \varphi^{(\ell)}\right) = \int\int \frac{1_{\{\bar{s}<t, \bar{s}<\bar{t}\}} \varphi^{(\ell)}(\bar{t}, x) \gamma_0^{(\ell)}(\bar{t})}{\left[1 - H^{(\ell)}(\bar{s})\right]^2} H_0^{(\ell)}(d\bar{s})\, H_{11}^{(\ell)}(\bar{t}, x),$$

where $H_0^{(\ell)}(t) = \mathbb{P}(Y \leq t, \delta = 0 | D = \ell)$.

Proposition 1.2 *Assume that $\mathbb{E}\left(X^{(\ell)} X^{(\ell)\prime}\right)$ is positive semidefinite. Under Assumptions 1.2-1.3 and Assumption 1.5 (in Appendix 1.7.1), for $\ell = \{0, 1\}$:*

$$n_\ell^{1/2}\left[\left(\hat{\beta}_\ell - \beta_\ell\right), \left(\hat{\mu}_X^{(\ell)} - \mu_X^{(\ell)}\right)\right] \xrightarrow{d} \mathcal{N}_{2k}\left(0, \Sigma^{(\ell)}_{\beta\mu_X}\right)$$

where

$$\Sigma^{(\ell)}_{\beta\mu_X} = \left(\Sigma^{(\ell)}_{XX}\right)^{-1} \Sigma^{(\ell)}_0 \left(\Sigma^{(\ell)}_{XX}\right)^{-1} = \left(\boldsymbol{\sigma}^{(\ell)}_\beta, \boldsymbol{\sigma}^{(\ell)}_{\beta\mu_X}; \boldsymbol{\sigma}^{(\ell)}_{\beta\mu_X}, \boldsymbol{\sigma}^{(\ell)}_{\mu_X}\right)$$

$$\Sigma^{(\ell)}_{XX} = \begin{pmatrix} \mathbb{E}\left(X^{(\ell)} X^{(\ell)\prime}\right) & 0 \\ 0 & I_k^{(\ell)} \end{pmatrix} \quad \text{and} \quad \Sigma^{(\ell)}_0 = \begin{pmatrix} \boldsymbol{\sigma}^{(\ell)}_{11} & \\ \boldsymbol{\sigma}^{(\ell)}_{12} & \boldsymbol{\sigma}^{(\ell)}_{22} \end{pmatrix}$$

and

$$\boldsymbol{\sigma}^{(\ell)}_{ij} = \mathbb{E}\left[\boldsymbol{\varphi}_i\left(Y^{(\ell)}, X^{(\ell)}\right) \boldsymbol{\varphi}_j\left(Y^{(\ell)}, X^{(\ell)}\right) \left(\gamma_0^{(\ell)}\left(Y^{(\ell)}\right)\right)^2 \delta^{(\ell)} - \gamma_1^{(\ell)}\left(Y^{(\ell)}; \boldsymbol{\varphi}_i\right) \gamma_1^{(\ell)}\left(Y^{(\ell)}; \boldsymbol{\varphi}_j\right) \left(1 - \delta^{(\ell)}\right)\right]$$

$\boldsymbol{\varphi}_i(t, x) = (\varphi_{i1}, \ldots, \varphi_{ik})$,

[8] Efron (1967) proposed an intuitive solution to reduce the bias by setting $\delta^{(\ell)}_{[n_\ell:n_\ell]} = 1$. Further discussion can be found in Meier (1975) and Mauro (1985).

$\varphi_{1l}(t,x) = x_l(t-\beta' x),$
$\varphi_{2l}(t,x) = x_l - \mu_X \quad \textit{for } 1 \leq l \leq k.$

Corollary 1.2 *Under Assumption 1.1 and the same conditions as in Proposition 1.2, and* $\frac{n_\ell}{n} \to \rho_\ell$ *with* $\rho_0 + \rho_1 = 1$*, we have:*

$$n^{1/2}\left(\hat{\Delta}_T^\mu - \Delta_T^\mu\right) \xrightarrow{d} \mathcal{N}(0, V_{\Delta_T})$$
$$n^{1/2}\left(\hat{\Delta}_S^\mu - \Delta_S^\mu\right) \xrightarrow{d} \mathcal{N}(0, V_{\Delta_S})$$
$$n^{1/2}\left(\hat{\Delta}_C^\mu - \Delta_C^\mu\right) \xrightarrow{d} \mathcal{N}(0, V_{\Delta_C})$$

where V_{Δ_T}*,* V_{Δ_S} *and* V_{Δ_C} *are defined in Appendix 1.7.1.*

Accordingly, a confidence interval of $100(1-2\alpha)\%$ for the structure effect and the composition effect are given by

$$\hat{\Delta}_S^\mu \pm \mathcal{Z}_{1-\alpha}\frac{\hat{V}_{\Delta_S}}{n^{1/2}} \quad \text{and} \quad \hat{\Delta}_C^\mu \pm \mathcal{Z}_{1-\alpha}\frac{\hat{V}_{\Delta_C}}{n^{1/2}}$$

where,

$$\hat{V}_{\Delta_S} = \frac{1}{1-\rho}\hat{\Delta}_\beta' \hat{\boldsymbol{\sigma}}_{\mu_X}^{(1)} \hat{\Delta}_\beta + \frac{2}{1-\rho}\hat{\Delta}_\beta' \hat{\boldsymbol{\sigma}}_{\beta\mu_X}^{(1)} \hat{\mu}_X^{(1)} + \frac{1}{\rho(1-\rho)}\hat{\mu}_X^{(1)\prime} \hat{\boldsymbol{\sigma}}_\beta \hat{\mu}_X^{(1)},$$
$$\hat{V}_{\Delta_C} = \frac{1}{\rho}\hat{\Delta}_{\mu_X}' \hat{\boldsymbol{\sigma}}_\beta^{(0)} \hat{\Delta}_{\mu_X} + \frac{2}{\rho}\hat{\beta}_0' \hat{\boldsymbol{\sigma}}_{\beta\mu_X}^{(0)} \hat{\Delta}_{\mu_X} + \frac{1}{\rho(1-\rho)}\hat{\beta}_0' \hat{\boldsymbol{\sigma}}_{\mu_X} \hat{\beta}_0,$$

$\mathcal{Z}_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution, $\hat{\Delta}_\beta = \hat{\beta}_1 - \hat{\beta}_0$, $\hat{\boldsymbol{\sigma}}_\beta = \rho\hat{\boldsymbol{\sigma}}_\beta^{(1)} + (1-\rho)\hat{\boldsymbol{\sigma}}_\beta^{(0)}$, $\hat{\Delta}_{\mu_X} = \hat{\mu}_X^{(1)} - \hat{\mu}_X^{(0)}$, $\hat{\boldsymbol{\sigma}}_{\mu_X} = \rho\hat{\boldsymbol{\sigma}}_{\mu_X}^{(1)} + (1-\rho)\hat{\boldsymbol{\sigma}}_{\mu_X}^{(0)}$ and $\rho_0 = \rho$. In addition, $\hat{\boldsymbol{\sigma}}_\beta^{(\ell)}$ and $\hat{\boldsymbol{\sigma}}_{\mu_X}^{(\ell)}$ are the empirical analog of $\boldsymbol{\sigma}_\beta^{(\ell)}$ and $\boldsymbol{\sigma}_{\mu_X}^{(\ell)}$ defined in Proposition 1.2. In the absence of censoring, these results coincide with those proposed by Jann (2005, 2008).

In general, computing the asymptotic variance of the decomposition components is cumbersome (c.f. Fortin et. al., 2011; Rothe, 2012), even more in the case of censoring when estimating $\gamma_0^{(\ell)}$ and $\gamma_1^{(\ell)}$ is needed. A practical alternative widely used in the decomposition methods literature, is the implementation of nonparametric bootstrap techniques. Proper resampling methods for censored data are briefly described in Section 1.6.

1.3 Decomposition based on Model Specification

When the mean difference is not informative to make comparisons between subpopulations, the counterfactural decomposition of other distributional features, such as the variance or the Gini coefficient, is compelling. Construct a model for a particular parameter could be costly if we are interested in more than one feature of the subpopulations. Instead, we can take into account a decomposition method based on the whole distribution and its associated functionals. In order to exploit the dependence between T and X to compute the composition effect and the structure effect, consider the counterfactual distribution of the subpopulation i given the characteristics of subpopulation j, say $F_T^{(i,j)}$, which can be defined in terms of the conditional distribution

$$F^{(\ell)}(t|x) = \mathbb{P}(T \leq t, D = \ell | X), \ \ell = \{0,1\},$$

then,

$$F_T^{(i,j)}(t) = \mathbb{E}\left[F^{(i)}(t|X)\,|D = j\right] = \int F^{(i)}(t|x)\, dF^{(j)}(x)$$

with $F^{(\ell)}(x) = \mathbb{P}(X \leq x | D = \ell)$, $\ell = \{0,1\}$.

The validity of this *counterfactual operator* (also discussed by Rothe, 2010; Chernozhukov et. al., 2013 and Donald and Hsu, 2014) follows the conditions stated in Assumption 1.1. That is, when varying the covariates distribution, the conditional distribution of unobservables is not affected, and the counterfactual distribution $F_T^{(i,j)}$ is defined as the integral of $F^{(i)}(t|x)$ over the covariates distribution of subpopulation j. If $F_T^{(i,j)}$ is identifiable, a large class of parameters

$\theta\left(F_T^{(i,j)}\right)$ can be decomposed as

$$\Delta_T^\theta = \theta\left(F_T^{(1,1)}\right) - \theta\left(F_T^{(0,1)}\right) + \theta\left(F_T^{(0,1)}\right) - \theta\left(F_T^{(0,0)}\right) = \Delta_S^\theta + \Delta_C^\theta \tag{1.7}$$

where Δ_S^θ and Δ_C^θ are the corresponding structure effect and composition effect. An interesting feature of this approach lies on that decompose parameters with complicated forms, like truncated moments or inequality measures, can be computed without any linear approximation.

An estimator of the counterfactual distribution $\bar{F}_T^{(i,j)}$ can be obtained by plugging-in the empirical analog of the multivariate distribution of covariates,

$$\bar{F}^{(\ell)}(x) = n_\ell^{-1} \sum_{i=1}^n 1_{\{X_i \leq x, D_i = \ell\}}$$

i.e., we have

$$\bar{F}_T^{(i,j)}(t) = n_j^{-1} \sum_{l=1}^n \bar{F}^{(i)}(t|x_l)\, 1_{\{D_l = j\}}.$$

As a consequence, the estimation procedure reduces to the identification and estimation of $F^{(\ell)}(t|x)$. Under Assumption 1.2, the conditional distribution $F^{(\ell)}(t|x)$ can be identified from (Y, X, δ, D); however, this assumption can be replaced by a weaker condition (Assumption 1.4), at the cost of imposing restrictions on the functional form.

Assumption 1.4 *For each $\ell = \{0, 1\}$, it holds that $Y^{(\ell)} \perp C^{(\ell)} | X^{(\ell)}$.*

This assumption has been taken into consideration to propose numerous generalizations of the Kaplan-Meier estimator (c.f. Beran, 1981; Dabrowska, 1987, 1989; Gonzalez-Manteiga and Cadarso-Suarez, 1994; Akritas, 1994; Leconte et. al., 2002), and it is more apropiate when strong independence fails. For instance, in presence of competing risk, individuals have more than one possible state destination and the likelihood of transition from one state to other might be correlated with their characteristics.

Because of the presence of censoring, classical methods such as quantile regression and distribution regression (c.f. Chernozhukov et. al., 2013; Koenker et. al., 2013), are not valid to estimate $F^{(\ell)}(t|x)$. There are quantile regression methods available for censored data (c.f. Ying et. al., 1995; Lipsitz et. al., 1997; Bang and Tsiatis, 2002; Portnoy, 2003; Peng and Huang, 2008; Wang and Wang, 2009; Gorfine et. al., 2014), but these procedures are computationally very demanding. Additionally, estimation involve to deal with nonsmooth and nonconvex objective functions, to make approximations around the tails of the distribution, and to carry out arrangements to guarantee monotonicity. Besides, the use of nonparametric models might be complicated since it is usual to deal with a large number of covariates, and establishing the limit process of the counterfactual operator would require further theoretical work (Rothe, 2010).

Alternatively, the conditional distribution can be recovered by exploiting the relation between F and Λ (see Equation –1.3–). So, we consider the popular proportional hazard specification proposed by Cox (1972, 1975) which assumes the following conditional hazard function:

$$\lambda^{(\ell)}(t|x) = \lambda_0^{(\ell)}(t)\,\phi(x, \beta_\ell) \tag{1.8}$$

where $\lambda_0^{(\ell)}$ is the baseline hazard (common risk) depending only on t and ϕ is a positive function representing the effect of the covariates on the conditional hazard function, commonly specified as $\phi(x, \beta_\ell) = \exp(\beta_\ell' x)$. In this context a fully parametric model is possible, but these models usually force the hazard function to be monotone. Instead, the Cox model does not require any shape assumption on $\lambda_0^{(\ell)}$. Moreover, this model is flexible for incorporating time-varying covariates and unobservable heterogeneity.

As a consequence of Equation (1.8), the conditional cumulative hazard is

$$\Lambda_T^{(\ell)}(t|x) = \int_0^t \lambda^{(\ell)}(\bar{t}|x)\, d\bar{t}$$

and hence, the conditional distribution is given by:

$$F^{(\ell)}(t|x) = 1 - \exp(-\Lambda_T^{(\ell)}(t|x)) = 1 - \left[\exp(-\Lambda_0^{(\ell)}(t))\right]^{\exp(\beta_\ell' x)}$$

where $\Lambda_0^{(\ell)}(t) = \int_0^y \lambda_0^{(\ell)}(\bar{t})\, d\bar{t}$. In order to estimate $F^{(\ell)}(t|x)$, Cox (1975) proposed the partial likelihood method which directly estimates β_ℓ and allows the nonparametric component $\Lambda_0^{(\ell)}(t)$ to be estimated.

With respect to the latter, there are two popular estimators in the literature. The first, and the most commonly used, is the Breslow estimator, $\hat{\Lambda}_{0B}^{(\ell)}$, introduced by Breslow (1974). This is given by

$$\hat{\Lambda}_{0B}^{(\ell)}(t) = \sum_{i=1}^{t} \frac{1}{\sum_{j \in r^{(\ell)}(t_i)} e^{\hat{\beta}_\ell' x_j^{(\ell)}}}.$$

with $r^{(\ell)}(t_i)$ the pool risk in subpopulation ℓ at period t_i and $d^{(\ell)}(t_i)$ the set of individuals in subpopulation ℓ changing state at period t_i. In turn, the second was proposed by Kalbfleisch and Prentice (1973), which constructs a discrete cumulative hazard that is consistent with the first order condition (or score) of the partial likelihood function, i.e.:

$$\hat{\Lambda}_{0KP}^{(\ell)}(t) = \sum_{i=1}^{n_\ell} \left(1 - \hat{\alpha}_i^{(\ell)}\right) 1_{\{t_i \leq t\}}$$

where the hazard probabilities $\hat{\alpha}_i^{(\ell)}$ solve:

$$\sum_{j \in d^{(\ell)}(t_i)} e^{\hat{\beta}_\ell' x_j^{(\ell)}} \left(1 - \hat{\alpha}_i^{\exp(\hat{\beta}_\ell' x_j^{(\ell)})}\right)^{-1} = \sum_{l \in r^{(\ell)}(t_i)} e^{\hat{\beta}_\ell' x_l^{(\ell)}}$$

$\hat{\Lambda}_{0B}^{(\ell)}$ presents practical advantages since it does not involve solving auxiliary equations. Both estimators perform similarly in finite sample, as will be discussed in Section 1.4.

Thus, an estimator for the counterfactual components of the decomposition in Equation (1.7), hereafter *Counterfactual Cox* decomposition (CCOX), is given by:

$$\hat{\Delta}^{\theta} = \theta\left(\hat{F}_T^{(1,1)}\right) - \theta\left(\hat{F}_T^{(0,1)}\right) + \theta\left(\hat{F}_T^{(0,1)}\right) - \theta\left(\hat{F}_T^{(0,0)}\right) \tag{1.9}$$

where

$$\hat{F}_T^{(i,j)}(t) = n_j^{-1}\sum_{l=1}^{n_j}\hat{F}^{(i)}(t|x_l)$$

and

$$\hat{F}^{(\ell)}(t|x) = 1 - \left[\exp(-\hat{\Lambda}_0^{(\ell)}(t))\right]^{\exp(\hat{\beta}_\ell' x)}. \tag{1.10}$$

The validity of the CCOX is established in Proposition 1.3.

Proposition 1.3 *Consider that Assumptions 1.1, 1.3 and 1.4, Assumption 1.6 (see Appendix 1.7.1) and Equation 1.8 hold, and* $\frac{n_\ell}{n} \to \rho_\ell$ *with* $\rho_0 + \rho_1 = 1$. *Then:*

$$n^{1/2}\left(\hat{F}_T^{(i,j)}(t) - F_T^{(i,j)}(t)\right) \Rightarrow \bar{M}^{(i,j)}(t)$$

where $\bar{M}_{ij}$ *is a tight zero-mean Gaussian process with uniform continuous path on* $Supp(T)$*, defined as:*

$$\bar{M}^{(i,j)}(t) = \rho_i^{1/2}\int M^{(i)}(t,x)\,dF^{(j)}(x) + \rho_j^{1/2}N^{(j)}\left(F^{(i)}(t|.)\right).$$

Moreover, since the limit process of $\hat{F}_T^{(i,j)}$ is nonpivotal (see Chernozhukov et. al., 2013), resampling methods are suitable for making inference on the counterfactual components (see Appendix 1.7.1 for further discussion).

1.4 Monte Carlo Simulations

To study finite sample properties of COB and CCOX, we carry out Monte Carlo experiments. These exercises allow the methods to be compared with other competing alternatives, and provide evidence on their performance under different censoring scenarios and distributional assumptions.

1.4.1 COB Decomposition

We study the performance of the COB procedure to estimate the structure and the composition effects with respect to two alternatives: the classical OB neglecting the presence of censoring, i.e. assuming that Y is the actual duration; and OB by dropping censored observations. To do so, we consider Data Generator Processes (DGPs) where T depends linearly on X and a vector or errors ε, as shown in Table 1.1. In addition, we assume a single covariate simulated as $X^{(0)} \sim \mathcal{N}(1.5, 0.5)$ and $X^{(1)} \sim \mathcal{N}(1, 0.5)$. In this scenario, $\Delta_S = 0$ and $\Delta_C = 0.5$. To adjust the censoring level to 30%, the mean of the error associated to C are shifted by $(v_0, v_1) = (2.5, 2)$. Finally, we consider sample size of 50, 500 and 2500 and evaluate the performance of $\hat{\Delta}_S$ and $\hat{\Delta}_C$ using a measure of bias, in particular, the average of absolute deviations across 1000 simulation draws.

Table 1.1 Simulation Setup COB method

$\ell = \mathbf{0}$	$T^{(0)} = 5 + X^{(0)} + \varepsilon_T^{(0)},\ \ \varepsilon_T^{(0)} \sim \mathcal{N}(0, 1)$
	$C^{(0)} = 5 + \varepsilon_C^{(0)},\ \ \varepsilon_C^{(0)} \sim \mathcal{N}(v_0, 1.5)$
$\ell = \mathbf{1}$	$T^{(1)} = 5 + X^{(1)} + \varepsilon_T^{(1)},\ \ \varepsilon_T^{(1)} \sim \mathcal{N}(0, 1)$
	$C^{(1)} = 5 + \varepsilon_C^{(1)},\ \ \varepsilon_C^{(1)} \sim \mathcal{N}(v_1, 1.5)$

Results in Table 1.6 show that there are important differences among alternative estimators when censoring is present, and most importantly, COB outperforms these alternatives. Furthermore, if censoring is ignored, the absolute bias is not reduced as sample size increases.

1.4.2 Censoring Mechanism and Distributional Assumption

In this exercise, we examine the performance of the counterfactual distribution operator based on the Cox model under different DGPs. This allows three relevant aspects to be studied: *i.* the censoring mechanism, *ii.* the distributional assumption on duration outcome, and *iii.* the estimator of the baseline cumulative hazard. The main parameters of the simulation are presented in Table 1.2. It is assumed a single covariate following a uniform distribution $\mathcal{U}(0,1)$. The scale and the shape of censoring times are shifted to generate censoring levels of 5%, 20% and 50% and sample sizes are set at 50, 500 and 2500.

Table 1.2 Simulation Setup Counterfactual Operator with Cox Model

Assumption		**DGP**
$T \perp C$	**Weibull**	$T \sim WB\left(e^{2-x}, 5\right)$ $C \sim WB\left(e^{2+\upsilon}, 5\right)$ $\upsilon = (0.25, -0.2, -0.5)$
	Normal	$T = 5 + X + \varepsilon_T,\ \ \varepsilon_T \sim N(0,1)$ $C = 5 + \varepsilon_C,\ \ \varepsilon_C \sim N(\upsilon, 1)$ $\upsilon = (3, 1.5, 0.5)$
$T \perp C \vert X$	**Weibull**	$T \sim WB\left(e^{2-x}, 5\right)$ $C \sim WB\left(e^{2-x+\upsilon}, 7\right)$ $\upsilon = (0.45, 0.2, -0.02)$
	Normal	$T = 5 + X + \varepsilon_T,\ \ \varepsilon_T \sim N(0,1)$ $C = 5 + X + \varepsilon_C,\ \ \varepsilon_C \sim N(\upsilon, 1)$ $\upsilon = (2.5, 1, 0)$

To evaluate the performance of the counterfactual operator, we compute the distance between the marginal distribution estimates using the CCOX estimator,

$\hat{F}_T^{(i,i)}(t)$, with respect to the marginal empirical distribution. In addition, to evaluate the sensitivity of the identification assumptions we also make this comparison for the Kaplan-Meier estimator (KM), which is the corresponding marginal distribution $\hat{F}^{(i)}(t,\infty)$. To do so, we compute three measures: the maximum distance (MD), the average distance (AD) and the mean squared distance (MSD). To be specific,

$$MD = \max_t \left| \tilde{F}_T(t) - \hat{F}_T(t) \right|, \quad AD = \frac{1}{n}\sum_{i=1}^{n} \left| \tilde{F}_T(t) - \hat{F}_T(t) \right|$$

$$MSD = \frac{1}{n}\sum_{i=1}^{n} \left(\tilde{F}_T(t) - \hat{F}_T(t) \right)^2,$$

where $\tilde{F}_T$ is the empirical distribution. For MD and AD we report the average over 1000 draws, while for the latter, the square root of the mean value.

Results in Table 1.7 suggest that, under the independence assumption, the KM estimator outperforms the CCOX estimator when censorship level is low, and there are not important differences with medium or heavy censoring levels (20% and 50%). In turn, under conditional independence, it is noticeable that the performance measures decrease faster for CCOX than KM with the sample size, and this fact is more remarkable as censoring becomes more substantial. This stresses on the relevance of the assumptions to validate the use of a particular method.

Regarding the distributional assumptions (see Tables 1.8 and 1.9), we can observe that the CCOX estimator performs fairly well even if survival times follow a normal distribution. With respect to the estimators of the baseline cumulative hazard, results are roughly the same, except for a very small sample where $\hat{\Lambda}_{0KP}$ outperforms the $\hat{\Lambda}_{0B}$ estimator. This is explained by the nature of $\hat{\Lambda}_{0KP}$ since it is proposed in the context of discrete survival times.

1.4.3 Decomposition Exercise and Inference

To study the finite sample performance of the CCOX to compute counterfactual decompositions beyond the mean, we consider a simulation exercise where all the difference between the two subpopulations is due to the shape covariates distribution. In particular, we set $X^{(1)}$ as uniform $(0,1)$ and $X^{(0)}$ as the sum of three independent uniform distributions in the interval $\left(0, \frac{1}{3}\right)$. Hence, we decompose the truncated mean at 15 and the quartiles. We set $T^{(\ell)} \sim \mathcal{WB}\left(e^{3-X^{(\ell)}}, 5\right)$, and to generate censoring levels of 30% $C^{(\ell)} \sim \mathcal{WB}\left(e^{3.17-X^{(\ell)}}, 5\right)$, where $\mathcal{WB}$ denotes the Weibull distribution. We consider $n_\ell = 500$ and 1000 draws (Figure 1.1 shows a typical draw of this simulation exercise).

We estimate the decomposition given by Equation (1.9) using the Breslow estimator for the baseline cumulative hazard and test the hypothesis $\Delta_S^\theta = 0$ using 1000 bootstrap repetitions. The resampling procedure is executed using the *simple* method, and coverage intervals (at 95% and 90% confidence level) are constructed according to *percentile* and *hybrid* methods (see Section 1.6 for details). Results in Table 1.10 suggest that the coverage rate is close to its nominal value and the accuracy improves if the two subpopulations exhibit similar censoring levels. Regarding the confidence intervals, the percentile method tends to outperform the hybrid method, although the difference is quite small.

1.5 Unemployment Duration Gender Gaps in Spain

Spain is an interesting case to study unemployment gender gaps. First, Spain has experienced one of the highest unemployment rates among OECD countries in the recent decades. According to official statistics (OECD, 2013), the average unemployment rate in OECD countries was around 6.8% and in the US 5% for the period 1995-2005, while in Spain it was 14%. Moreover, the difference in unemployment rates by gender has also been important. For the same period, women exhibited on average an unemployment rate 9 percentage points (p.p.) higher, while in the US such a difference was rather than slight (0.04 p.p.).

There are a number of studies exploring the gender gaps in the unemployment rate (see for instance, Niemi, 1974; Johnson, 1983; Azmat et. al., 2006; Queneau and Sen, 2007), but other aspects of unemployment have been neglected. Therefore, we provide additional evidence of the unemployment gender gap in Spain by analyzing the duration rather than the rate. In particular, we estimate the total gender gap and perform counterfactual decomposition analysis to examine to what extent this gap is explained by the difference in observable workers' socioeconomic characteristics.

Literature devoted to studying unemployment duration gaps has focused exclusively in explaining the difference in the average conditional hazard rate (see Ham et. al., 1999, for the Czech and Slovak Republic; Gonzalo and Saarela, 2000, for Finland; Eusamio, 2004, for Spain and Portugal; Ortega, 2008, for Argentina; Du and Dong, 2009, 2009, for China; Tansel and Tasci, 2010, for Turkey; and Baussola et. al., 2015, for Italy and the UK). In these exercises interpretation of the decomposition components is difficult since the average conditional hazard rate does not correspond to the hazard rate. Instead, we use the proposed methods to decompose several parameters associated to the underlying unemployment duration distribution.

In particular, we study the gender gap in the average unemployment duration, the probability of being long term unemployed (12 and 24 months or longer) and the Gini coefficient. While the mean gives a broad picture of the difference in unemployment duration by gender, the other two parameters allow the difference in terms of severity to be analyzed. Considering a measure of inequality is interesting since, as in the case of income, unemployment duration has normative implications on social welfare (c.f. Paul, 1992; Borooah, 2002; Sengupta, 2009; Shorrocks, 2009a,b).

We explore two dimensions of unemployment duration: the duration until exit from unemployment and the duration until getting a job. To analyze the latter case, we follow the competing risk approach (similar to Addison and Portugal, 2003), by considering as censored all transitions to a destination other than getting a job. The distinction of the two types of duration is important because they have different policy implications; and also, it allows to illustrate the role of the identification assumptions related to the censoring mechanism. In the case of duration to exit from unemployment, censoring can be considered as administrative; but when the transition unemployment-to-employment is studied, censoring might not be independent since workers' characteristics affect the decision of being employed or out of the labor force.

To do so, we use information from the Survey of Income and Living Conditions (SILC) for the period 2004-2007. This survey, managed by the European Commission, is a rotative household panel that collects information on socioeconomic characteristics, including the occupational status (monthly) for a period of 4 years. Our population consists of unemployed workers older than 25 starting their unemployment spell during the period 2004-2007. We take into account a set of explanatory variables commonly used in unemployment duration studies such as age, educational level, tenure, marital status, whether the individual is head of the household, and the number of unemployed in the household (see for instance, Foley, 1997; Addison and Portugal, 2003; Kuhn and Skuterud, 2004; Biewen and Wilke, 2004; Tansel and Tasci, 2010). The first three variables control by human capital characteristics, while the rest give information about the opportunity cost of being unemployed and the reservation wage. In addition, we include city size and region to control for specific labor market characteristics.

In the case of the duration until leaving unemployment, the censoring levels are 21.4% for women and 16.2% for men. Based on the marginal distributions $\hat{F}^{(\ell)}(t,\infty)$ and $\hat{F}_T^{(\ell,\ell)}$, we compute the average duration, the probability of being long term unemployed (LTU) and the Gini coefficient (see Table 1.3). It is noticeable that estimates are very similar. To give some insight about the misleading conclusions produced by ignoring the censoring, the bottom part of Table 1.3 includes the estimates when censored observations are dropped. As expected, estimates are lower in the case of the average, and for the LTU and the Gini coefficient remarkable differences are also found.

Table 1.3 Distributional Parameters of Duration to Exit from Unemployment

		Mean	LTU(12)	LTU(24)	Gini
Kaplan-Meier integrals	**Women**	11.090	0.410	0.145	0.496
	Men	7.804	0.237	0.065	0.542
CCOX	**Women**	11.160	0.396	0.145	0.508
	Men	7.767	0.235	0.067	0.544
Only Uncensored	**Women**	7.456	0.292	0.045	0.446
	Men	5.466	0.153	0.014	0.485

Note: Authors' calculations. LTU: Long Term Unemployment.

Following the COB and CCOX methods, we compute the total difference and the decomposition components[9]. Results are presented in Table 1.4 coupled with

[9] In the case of the CCOX method, we check the validity of the proportional hazard assumption using the Schoenfeld residuals. The p-values (0.031 and 0.654 for women and men, respectively) suggest that in both cases the proportional hazard specification might be suitable.

confidence intervals at 90% built through 1000 bootstrap repetitions by using the percentile method. For the case of the average unemployment duration, results across methods are qualitatively the same and quantitatively similar. In general, it is observed that women present higher average duration and higher survival probability. In the case of the Gini coefficient, the difference is negative indicating that men's duration distribution is more unequal, which is consistent with the fact that men leave unemployment faster, on average, but they are also severely affected by LTU.

With respect to the decomposition factors, it is found that the structure effect is statistically different from zero and plays a major role in explaining the gender gap. Although not significant, the composition effect is always positive indicating that the difference in workers' characteristics slightly increases the severity of unemployment to the detriment of women. The structure effect is positive except in the case of the gender gap in the Gini coefficient, suggesting that factors others than workers' characteristics, i.e. institutional factors, labor market circumstances, behavioral aspects[10], among others, increase the average duration and probability of being LTU. Lastly, an interesting finding is that the gap in LTU(24) is lower than in LTU(12), and that such reduction is due to the decrease in the structure effect, implying that women are relatively less prone to experience long term unemployment, which agrees with the negative sign in the Gini coefficient difference.

In the second exercise, we study the gender gaps in unemployment duration until get a job[11]. Results of this decomposition are presented in Table 1.5. It can be observed that, in contrast to the previous exercise, the decomposition factors

[10] Indeed, this factor has been also named as behavioral effect (see c.f. Bachmann and Sinning, 2016).

[11] As before, we test the validity of the proportional hazard assumption, obtaining p-vaues of 0.283 and 0.410 for woman and men respectively. We also test the presence of unobservable heterogeneity at region level, but the hypothesis was rejected.

Table 1.4 Decomposition Distributional Statistics of Duration to Exit from Unemployment

			Total	Composition	Structure
COB	Mean	Difference	3.285	0.386	2.899
		CI 90%	[2.067 , 4.442]	[-0.478 , 2.425]	[0.408 , 4.219]
CCOX	Mean	Difference	3.392	0.537	2.855
		CI 90%	[2.098 , 4.491]	[-0.349 , 1.361]	[1.501 , 4.224]
	LTU(12)	Difference	0.161	0.012	0.148
		CI 90%	[0.115 , 0.206]	[-0.013 , 0.043]	[0.092 , 0.198]
	LTU(24)	Difference	0.078	0.014	0.064
		CI 90%	[0.043 , 0.109]	[-0.008 , 0.035]	[0.022 , 0.104]
	Gini	Difference	-0.036	0.006	-0.042
		CI 90%	[-0.071 , 0.000]	[-0.002 , 0.010]	[-0.075 , -0.005]

Note: Authors' calculations. Confidence interval computed by bootstrapping with 500 draws.

differ importantly between the COB and CCOX methods, which can be related to the validity of the identification assumptions. As mentioned previously, the independence assumption between survival times and censoring times might be strong to study the duration until getting a job. Despite the fact that the identification assumptions cannot be tested, we provide some suggestive evidence on the relation between the probability of censoring and the covariates.

Table 1.11 presents measures of goodness-of-fit estimated based on probability models for the censoring indicator on the covariates. In particular, we estimate linear probability models and logit models, and report the corresponding R^2-adjusted and *p-seudo* R^2. Overall, we observe that the covariates are relatively more important for predicting the censoring indicator for the duration until getting a job. Therefore, Assumption 1.2 might not be appropriate in this context. Intuitively, in absence of censoring, assumptions on censoring nature play no role and our methods should report similar results. So, we provide some additional evidence by performing the decomposition eliminating the censored observations (see Table

1.12), obtaining similar results for the two kinds of durations.

Hence, focusing on the CCOX estimates, results are qualitatively similar to the case of duration to leave unemployment. Moreover, we observe that the magnitudes of the total differences are higher, implying that inactivity is an important option for women, which is consistent with the argument that women are less attached to the labor market. This fact is also proven by the persistent difference in the LTU, i.e. the total difference in the probability of being unemployed in the long term does not decrease over duration spells.

Table 1.5 Decomposition Distributional Statistics of Duration from Unemployment to Employment

			Total	**Composition**	**Structure**
COB	**Mean**	**Difference**	7.224	5.698	1.525
		CI 90%	[4.804 , 9.020]	[3.816 , 8.678]	[-1.691 , 3.203]
CCOX	**Mean**	**Difference**	7.865	1.713	6.151
		CI 90%	[5.266 , 9.507]	[0.159 , 3.028]	[3.813 , 8.191]
	LTU(12)	**Difference**	0.184	0.036	0.148
		CI 90%	[0.137 , 0.231]	[0.000 , 0.071]	[0.095 , 0.202]
	LTU(24)	**Difference**	0.176	0.041	0.135
		CI 90%	[0.130 , 0.226]	[0.003 , 0.074]	[0.084 , 0.192]
	Gini	**Difference**	-0.040	-0.014	-0.025
		CI 90%	[-0.079 , -0.008]	[-0.029 , 0.001]	[-0.065 , 0.007]

Note: Authors' calculations. Confidence interval computed by bootstrapping with 500 draws.

Likewise, the composition effect turns out statistically significant, except for the Gini coefficient, and the structure effect has the most relevant role in explaining the unemployment duration gender gaps. This result has been also reported by Ham et. al. (1999); Gonzalo and Saarela (2000); Eusamio (2004) and Ortega (2008) who study the average conditional hazard rate. It is remarkable that this *unexplained* component is associated to many factors involved in the job search process, i.e. the behavior of workers and employers and different circumstances of the labor market such as the labor market tightness, and discrimination, among others. Thus, our results point out the importance of deeply studying such factors to assess the differential gender effect of labor market policies.

1.6 Final Remarks and Further Research

We have proposed inferential tools to perform counterfactual decompositions under right censoring, a common feature of duration outcomes. These tools encompass decompositions for the mean difference as well as the for other distributional features. For the mean difference, we provide a regression-based method and develop the asymptotic results useful to test statistically the significance of the decomposition components. In turn, the decomposition of other parameters is based on the estimation of the whole marginal distribution of counterfactual outcomes, which requires the specification of a hazard model to recover the conditional distribution of the duration outcome given the covariates. These methods have many potential applications to study duration outcomes of common interest in economics.

Although we focus in the aggregate decomposition, the COB decomposition method can be extended to perform detailed decomposition by addressing the usual issues such as path dependence (c.f. Firpo et. al., 2007; Firpo and Pinto, 2011; Rothe, 2012; Shorrocks, 2013) and omitted group problem (c.f. Oaxaca and Ransom, 1999; Gardeazabal and Ugidos, 2004; Yun, 2005). Also, the COB method can be adapted to perform counterfactual decompositions of other distributional features. To do so, the weighted regression method might be used to estimate the conditional recentered influence function (RIF) for several distributional statistics, as proposed in Firpo et. al. (2009), by applying the proper transformation in the dependent variable. A list of RIFs for relevant distributional parameters can be found in Firpo et. al. (2007) and Essama-Nssah and Lambert (2011).

In the context of counterfactual decompositions, bootstrapping methods turn out to be practical to make statistical inference. The implementation of statistical inference based on bootstrapping techniques is more accurate than first-order asymptotic approximation (Hall, 1992). Efron (1981) presents two alternative resampling schemes that have been recognized in the literature as the *simple* bootstrap method and the *obvious* bootstrap method. In short, the simple method consists in drawing bootstrap samples $(Y^*, X^*, D^*, \delta^*)$ by independent sampling of size n with replacement and assigning equal mass n^{-1} at each selected observation. Instead, the obvious method requires estimating the distribution of the survival times and censoring times. In particular, for each draw (X_i^*, D_i^*), compute $T_i^* \sim \hat{F}^{(\ell)}(t|x)$, and $C_i^* \sim \hat{G}^{(\ell)}(t|x)$ and define $Y_i^* = \min(T_i^*, C_i^*)$ and $\delta_i^* = 1_{\left\{T_i^* \leq C_i^*\right\}}$. Under independence between T and C, these methods are equivalent (c.f. Efron and Tibshirani, 1986). The simple method has important practical convenience because it does not require imposing any assumption on the structure of the data and does not depend on the censoring mechanism.

To construct confidence bands, we consider classical methods such as *percentile* and *hybrid* (see Hall, 1988; Efron, 1992; Burr, 1994, for a detailed comparison of coverage bands construction methods). One important advantage of these methods is that estimation of variances is not needed. In order to describe the pivotal quantities, suppose we are interested in forming $100(1-2\alpha)\%$ confidence bands for the target parameter θ. Denote the estimated parameter from a bootstrap sample as $\hat{\theta}^*$ and its distribution given by J. The percentile method sets the confidence interval as:

$$\left(J^{-1}(\alpha), \quad J^{-1}(1-\alpha)\right)$$

Instead of approximating the distribution of $\hat{\theta}^*$, the hybrid method approximates the distribution of $\left(\hat{\theta}-\theta\right)$ through the distribution of $\left(\hat{\theta}^*-\hat{\theta}\right)$. Therefore, the coverage interval is defined as:

$$\left(2\hat{\theta} - J^{-1}\left(1-\alpha\right),\quad 2\hat{\theta} - J^{-1}\left(\alpha\right)\right)$$

There is not a general rule to select the proper method. For instance, in the particular case of censored data, considering real-valued and function-valued parameters estimated through the Cox model, Burr (1994) makes comparative analysis of bootstrap confidence intervals combining resampling and interval construction methods. The results suggest that there is no single winner and the pertinence of each method depends on the target parameter.

1.7 Appendix

1.7.1 Some Theoretical Results

Proof of Proposition 1.1 and Estimation of the Joint Distribution

To achieve identification of the joint distribution we consider Assumption 1.2. Under Assumption 1.2.b. we have:

$$\begin{aligned}
H_{11}^{(\ell)}(t,x) &= \mathbb{P}(Y \le t, X \le x, \delta = 1 | D = \ell) \\
&= \mathbb{P}(\min(T,C) \le t, X \le x, \delta = 1 | D = \ell) \\
&= \mathbb{P}(T \le t, X \le x, T \le C | D = \ell) \\
&= \mathbb{E}\left[1_{\{T \le t\}} 1_{\{X \le x\}} \mathbb{P}(T \le C | T, X, D) | D = \ell\right] \\
&= \mathbb{E}\left[1_{\{T \le t\}} 1_{\{X \le x\}} \mathbb{P}(T \le C | T, D) | D = \ell\right] \\
&= \mathbb{E}\left[1_{\{T \le t\}} 1_{\{X \le x\}} \mathbb{P}(C \ge t | T, D) | D = \ell\right] \\
&= \mathbb{E}\left[1_{\{T \le t\}} 1_{\{X \le x\}} [1 - G(T- | D = \ell)] | D = \ell\right] \\
&= \int_0^t [1 - G(\bar{t} - | D = \ell)] F^{(\ell)}(d\bar{t}, x | D = \ell)
\end{aligned}$$

and by Assumption 1.2.a.

$$\begin{aligned}
1 - H^{(\ell)}(t) &= \mathbb{P}(Y > t | D = \ell) \\
&= \mathbb{P}(T > t, C > t | D = \ell) \\
&= \mathbb{P}(T > t | D = \ell) \mathbb{P}(C > t | D = \ell) \\
&= \left[1 - F_T^{(\ell)}(t)\right] [1 - G(t | D = \ell)] \\
&= \left[1 - F^{(\ell)}(t, \infty)\right] [1 - G(t | D = \ell)]
\end{aligned}$$

Thus, using Equation (1.3)

$$\begin{aligned}
\Lambda^{(\ell)}(t,x) &= \int_0^t \frac{F^{(\ell)}(d\bar{t},x)}{1-F^{(\ell)}(\bar{t}-,\infty)} \\
&= \int_0^t \frac{F^{(\ell)}(d\bar{t},x)\left[1-G(T-|D=\ell)\right]}{\left[1-F^{(\ell)}(\bar{t}-,\infty)\right]\left[1-G(T-|D=\ell)\right]} \\
&= \int_0^t \frac{H_{11}^{(\ell)}(d\bar{t},x)}{1-H^{(\ell)}(\bar{t}-)}.
\end{aligned}$$

Proof of Corollary 1.1

Define $S^{(\ell)}(t) = 1 - F_T^{(\ell)}(t)$. The joint distribution can be written as:

$$\begin{aligned}
F^{(\ell)}(t,x) &= \int_0^t F^{(\ell)}(d\bar{t},x) \\
&= \int_0^t \left[1-F_T^{(\ell)}(\bar{t}-)\right] \frac{F^{(\ell)}(d\bar{t},x)}{\left[1-F_T^{(\ell)}(\bar{t}-)\right]} \\
&= \int_0^t S^{(\ell)}(\bar{t})\, \Lambda^{(\ell)}(d\bar{t},x) \\
&= \int_0^t S^{(\ell)}(\bar{t})\, \frac{H_{11}^{(\ell)}(d\bar{t},x)}{\left[1-H^{(\ell)}(\bar{t}-)\right]}.
\end{aligned}$$

Using the sample version of $H_{11}^{(\ell)}$ and $H^{(\ell)}$ given by

$$\hat{H}^{(\ell)}(t) = n_\ell^{-1}\sum_{i=1}^{n} 1_{\{Y_i \le t, D_i=\ell\}} \quad and \quad \hat{H}_{11}^{(\ell)}(t,x) = n_\ell^{-1}\sum_{i=1}^{n} 1_{\{Y_i \le t, X_i \le x, \delta_i=1,\ D_i=\ell\}}$$

the jump $\hat{\Lambda}^{(\ell)}$ is defined as

$$\hat{\Lambda}^{(\ell)}(\{\bar{t}\},x) = \frac{\delta_{[i:n_\ell]}^{(\ell)}}{n_\ell - R_i^{(\ell)} + 1}.$$

Therefore, $\hat{F}^{(\ell)}(t,x)$ can be estimated by:

$$\hat{F}^{(\ell)}(t,x) = 1 - \prod_{\bar{t}\le t}\left[1-\hat{\Lambda}^{(\ell)}\left(\{\bar{t}\},x\right)\right] = 1 - \prod_{Y^{(\ell)}_{i:n_\ell}\le t, X^{(\ell)}_{[i:n_\ell]}\le x}\left[1-\frac{\delta^{(\ell)}_{[i:n_\ell]}}{n_\ell - R^{(\ell)}_i + 1}\right].$$

Additionally, $S^{(\ell)}$ is consistently estimated by the Kaplan-Meier estimator (by Assumption 1.2.a.). Thus,

$$\hat{F}^{(\ell)}(t,x) = \int_0^t \hat{S}^{(\ell)}(\bar{t})\,\frac{\hat{H}^{(\ell)}_{11}(d\bar{t},x)}{\left[1-\hat{H}^{(\ell)}(\bar{t}-)\right]}$$

where,

$$\hat{S}^{(\ell)}(t) = \prod_{Y^{(\ell)}_{i:n_\ell}\le t}\left[1-\frac{\delta^{(\ell)}_{[i:n_\ell]}}{n_\ell - R^{(\ell)}_i + 1}\right].$$

Thus, an estimator of the joint distribution is given by:

$$\begin{aligned}
\hat{F}^{(\ell)}(t,x) &= \sum_{i=1}^{n}\prod_{j=1}^{i-1}\left[1-\frac{\delta^{(\ell)}_{[j:n_\ell]}}{n_\ell - R^{(\ell)}_j + 1}\right]\frac{1_{\left\{Y^{(\ell)}_{i:n_\ell}\le t, X^{(\ell)}_{[i:n_\ell]}\le x, \delta^{(\ell)}_{[i:n_\ell]}=1\right\}}}{n_\ell - R^{(\ell)}_i + 1}\\
&= \sum_{i=1}^{n}\left\{\prod_{j=1}^{i-1}\left[1-\frac{\delta^{(\ell)}_{[j:n_\ell]}}{n_\ell - R^{(\ell)}_j + 1}\right]\frac{\delta^{(\ell)}_{[i:n_\ell]}}{n - R^{(\ell)}_i + 1}\right\}1_{\left\{Y^{(\ell)}_{i:n_\ell}\le t, X^{(\ell)}_{[i:n_\ell]}\le x\right\}}\\
&= \sum_{i=1}^{n} W^{(\ell)}_i 1_{\left\{Y^{(\ell)}_{i:n_\ell}\le t, X^{(\ell)}_{[i:n_\ell]}\le x\right\}}
\end{aligned}$$

where

$$W^{(\ell)}_i = \frac{\delta^{(\ell)}_{[i:n_\ell]}}{n_\ell - R^{(\ell)}_i + 1}\prod_{j=1}^{i-1}\left[1-\frac{\delta^{(\ell)}_{[j:n_\ell]}}{n_\ell - R^{(\ell)}_j + 1}\right].$$

Proof of Proposition 1.2

Before state the main result, we require some regularity conditions on the generic integrable functions $\varphi^{(\ell)}$.

Assumption 1.5 *Consider that following integrability conditions holds:*
a. $\int \left[\varphi^{(\ell)}(T,X)\gamma_0^{(\ell)}(T)\,\delta^{(\ell)}\right] d\mathbb{P} < \infty$ *.*
b. $\int \left|\varphi^{(\ell)}(\bar{t},\bar{x})\right| K^{1/2}(\bar{t},\ell)\, d\mathbb{P} < \infty$ *with*

$$K^{(\ell)}(t) = \int_0^{t-} \frac{G^{(\ell)}(d\bar{t})}{[1-H^{(\ell)}(\bar{t})]\,[1-G^{(\ell)}(\bar{t})]}.$$

Condition 1.5.a. generalizes the second order assumption on $\varphi^{(\ell)}$ so that when censoring is not present, this condition states that the second moment is finite. In turn, Condition 1.5.b controls the bias of $\mathbb{E}\left[\varphi^{(\ell)}(t,x)\right]$ and guarantees that censoring effects does not dominate in the right tail.

Now define:

$$Q_{XX}^{(\ell)} = \sum_{i=1}^{n_\ell} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)} X_{[i:n_\ell]}^{(\ell)\prime}$$

By Theorem 1 in Stute (1993), we have $Q_{XX}^{(\ell)} \longrightarrow \mathbb{E}\left[X^{(\ell)}X^{(\ell)\prime}\right]$.

To compute the joint distribution of $\left[\left(\hat{\beta}_\ell - \beta_\ell\right), \left(\hat{\mu}_X^{(\ell)} - \mu_X^{(\ell)}\right)\right]$, note that we can write:

$$\hat{\beta}_\ell = Q_{XX}^{(\ell)^{-1}} \sum_{i=1}^{n_\ell} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)} Y_{i:n_\ell}^{(\ell)}.$$

For all i, we know that $\delta_{[i:n_\ell]}^{(\ell)} Y_{i:n_\ell}^{(\ell)} = \delta_{[i:n_\ell]}^{(\ell)} T_{i:n_\ell}^{(\ell)}$. Then, we have:

$$\hat{\beta}_\ell - \beta_\ell = Q_{XX}^{(\ell)^{-1}} \sum_{i=1}^{n_\ell} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)} \varepsilon_{i:n_\ell}^{(\ell)}$$

Therefore,

$$
\begin{aligned}
\begin{pmatrix} \hat{\beta}_\ell - \beta_\ell \\ \hat{\mu}_X^{(\ell)} - \mu_X^{(\ell)} \end{pmatrix} &= \begin{pmatrix} Q_{XX}^{(\ell)^{-1}} & 0 \\ 0 & I_K \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)} \varepsilon_{[i:n_\ell]} \\ \sum_{i=1}^n W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)} - \mu_X^{(\ell)} \end{pmatrix} \\
&= \boldsymbol{Q}_{XX}^{(\ell)^{-1}} \begin{pmatrix} \sum_{i=1}^n W_i^{(\ell)} \boldsymbol{\varphi}_1^{(\ell)}(Y_i^{(\ell)}, X_i^{(\ell)}) \\ \sum_{i=1}^n W_i^{(\ell)} \boldsymbol{\varphi}_2^{(\ell)}(Y_i^{(\ell)}, X_i^{(\ell)}) \end{pmatrix} \\
&= \boldsymbol{Q}_{XX}^{(\ell)^{-1}} \boldsymbol{U}
\end{aligned}
$$

where:

- $\boldsymbol{\varphi}_i\,(t,x) = (\varphi_{i1}, \ldots, \varphi_{ik})$
- $\varphi_{1l}\,(t,x) = x_l\,(t - \beta' x) \quad \text{for } 1 \leq l \leq k$
- $\varphi_{2l}\,(t,x) = x_l - \mu_X^{(\ell)} \quad \text{for } 1 \leq l \leq k.$

From the SLLN (Stute, 1993), it follows that $\boldsymbol{U} \longrightarrow \mathbf{0}$. For a generic integrable $\varphi^{(\ell)}$, under Condition 1.5, $Q_\varphi^{(\ell)} = \int \varphi^{(\ell)}(t,x)dF$ admits the following representation (see Stute, 1996 for details):

$$
\begin{aligned}
Q_\varphi^{(\ell)} &= \sum_{i=1}^{n_\ell} W_i^{(\ell)} \varphi(Y_{i:n_\ell}^{(\ell)}, X_{[i:n_\ell]}^{(\ell)}) \\
&= \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} \Big[\varphi^{(\ell)}(Y_i^{(\ell)}, X_i^{(\ell)}) \gamma_0^{(\ell)} \left(Y_i^{(\ell)}\right) \delta_i^{(\ell)} + \gamma_1^{(\ell)} \left(Y_i^{(\ell)}; \varphi^{(\ell)}\right) \left(1 - \delta_i^{(\ell)}\right) \\
&\quad - \gamma_2^{(\ell)} \left(Y_i^{(\ell)}; \varphi^{(\ell)}\right) \Big] + o_{\mathbb{P}} \left(n_\ell^{-1/2}\right) \\
&= \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} \eta_i^{(\ell)} \left(Y_i^{(\ell)}, X_i^{(\ell)}; \varphi^{(\ell)}\right) + o_{\mathbb{P}} \left(n_\ell^{-1/2}\right)
\end{aligned}
$$

with

$$\eta_i^{(\ell)}\left(Y_i^{(\ell)},X_i^{(\ell)};\varphi^{(\ell)}\right) = \varphi^{(\ell)}(Y_i^{(\ell)},X_i^{(\ell)})\gamma_0^{(\ell)}\left(Y_i^{(\ell)}\right)\delta_i^{(\ell)} + \gamma_1^{(\ell)}\left(Y_i^{(\ell)};\varphi^{(\ell)}\right)\left(1-\delta_i^{(\ell)}\right) - \gamma_2^{(\ell)}\left(Y_i^{(\ell)};\varphi^{(\ell)}\right) = A_i^{(\ell)} + B_i^{(\ell)} - C_i^{(\ell)} \tag{1.11}$$

a sum of *iid* quantities such that:

- $\mathbb{E}\left[A_i^{(\ell)}\right] = \mathbb{E}\left[\varphi^{(\ell)}(Y_i^{(\ell)},X_i^{(\ell)})\right]$
- $\mathbb{E}\left[B_i^{(\ell)}\right] = \mathbb{E}\left[C_i^{(\ell)}\right]$

In such manner, we have the following result:

$$n_\ell^{1/2}\boldsymbol{U} \xrightarrow{d} \mathcal{N}_{2k}\left(0,\Sigma_0^{(\ell)}\right)$$

where

$$\Sigma^{(\ell)} = \begin{pmatrix} \boldsymbol{\sigma}_{11}^{(\ell)} & \cdot \\ \boldsymbol{\sigma}_{12}^{(\ell)} & \boldsymbol{\sigma}_{22}^{(\ell)} \end{pmatrix}$$

$$\boldsymbol{\sigma}_{ij}^{(\ell)} = \mathbb{C}ov\left[\boldsymbol{\eta}^{(\ell)}\left(Y^{(\ell)},X^{(\ell)};\boldsymbol{\varphi}_i^{(\ell)}\right),\boldsymbol{\eta}^{(\ell)}\left(Y^{(\ell)},X^{(\ell)};\boldsymbol{\varphi}_j^{(\ell)}\right)\right]$$

where each element of the vector $\boldsymbol{\eta}^{(\ell)}$ can be written as in Equation (1.11).

To simplify notation, let's omit "$^{(\ell)}$".Since each component of $\boldsymbol{\eta}$ has zero mean, the covariance can be written as:

$$\begin{aligned}\mathbb{C}ov\left[\boldsymbol{\eta}\left(Y,X;\boldsymbol{\varphi}_i\right),\boldsymbol{\eta}\left(Y,X;\boldsymbol{\varphi}_j\right)\right] &= \mathbb{E}\left(\boldsymbol{A}_i\boldsymbol{A}_j\right)+\mathbb{E}\left(\boldsymbol{A}_i\boldsymbol{B}_j\right)-\mathbb{E}\left(\boldsymbol{A}_i\boldsymbol{C}_j\right)\\ &+\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{A}_j\right)+\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{B}_j\right)-\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{C}_j\right)\\ &-\mathbb{E}\left(\boldsymbol{C}_i\boldsymbol{A}_j\right)-\mathbb{E}\left(\boldsymbol{C}_i\boldsymbol{B}_j\right)+\mathbb{E}\left(\boldsymbol{C}_i\boldsymbol{C}_j\right)\end{aligned}$$

Azarang at. al. (2015) shown that:

- $\mathbb{E}\left(\boldsymbol{A}_i \boldsymbol{C}_j\right) = \mathbb{E}\left(\boldsymbol{B}_i \boldsymbol{B}_j\right)$
- $\mathbb{E}\left(\boldsymbol{C}_i \boldsymbol{C}_j\right) = \mathbb{E}\left(\boldsymbol{B}_i \boldsymbol{C}_j\right) + \mathbb{E}\left(\boldsymbol{C}_i \boldsymbol{B}_j\right)$

In addition, $\mathbb{E}\left(\boldsymbol{A}_i \boldsymbol{B}_j\right) = \mathbb{E}\left(\boldsymbol{B}_i \boldsymbol{A}_j\right) = 0$ so that the covariance becomes:

$$\begin{aligned} \boldsymbol{\sigma}_{ij} &= \mathbb{E}\left(\boldsymbol{A}_i \boldsymbol{A}_j\right) - \mathbb{E}\left(\boldsymbol{B}_i \boldsymbol{B}_j\right) \\ &= \mathbb{E}\left[\boldsymbol{\varphi}_i\left(Y,X\right)\boldsymbol{\varphi}_j\left(Y,X\right)\left(\gamma_0\left(Y\right)\right)^2\delta - \gamma_1\left(Y;\boldsymbol{\varphi}_i\right)\gamma_1\left(Y;\boldsymbol{\varphi}_j\right)\left(1-\delta\right)\right] \end{aligned}$$

Finally, as

$$\boldsymbol{Q}_{XX}^{(\ell)} \longrightarrow \Sigma_{XX}^{(\ell)} = \begin{pmatrix} \mathbb{E}\left(XX'\right) & 0 \\ 0 & I_K \end{pmatrix}$$

we get our result.

Note that an estimator of the variance is obtained by plugging-in the sample analogs of $\Sigma_{XX}^{(\ell)}$, $\gamma_0^{(\ell)}$ and $\gamma_1^{(\ell)}$. Moreover, in absence of censoring, $\gamma_0^{(\ell)} = \delta^{(\ell)} = 1$ and it arrives to the classical result given by:

$$\boldsymbol{\sigma}_{ij}^{(\ell)} = \mathbb{E}\left[\boldsymbol{\varphi}_i^{(\ell)}\left(Y,X\right)\boldsymbol{\varphi}_j^{(\ell)}\left(Y,X\right)\right]$$

where $\boldsymbol{\sigma}_{12}^{(\ell)} = 0$ as long as $\mathbb{E}\left(\varepsilon|X\right) = 0$ holds.

Proof of Corollary 1.2

By Proporsition 1.2, we can write:

$$n^{1/2}\begin{pmatrix} \hat{\beta}_\ell - \beta_\ell \\ \hat{\mu}_X^{(\ell)} - \mu_X^{(\ell)} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\rho_\ell}\Sigma_{\beta\mu_X}^{(\ell)}\right)$$

where $\frac{n_\ell}{n} \to \rho_\ell$.

Then,

$$n^{1/2}\left(\hat{\beta}_\ell'\hat{\mu}_X^{(\ell)} - \beta_\ell'\mu_X^{(\ell)}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\rho_\ell}V_\ell\right)$$

with

$$V_\ell = \mu_X^{(\ell)\prime}\boldsymbol{\sigma}_\beta^{(\ell)}\mu_X^{(\ell)} + \beta_\ell'\boldsymbol{\sigma}_{\mu_X}^{(\ell)}\beta_\ell + 2\beta_\ell'\boldsymbol{\sigma}_{\beta\mu_X}^{(\ell)}\mu_X^{(\ell)}$$

Denote $\rho_0 = \rho$. Thus, for the total mean difference, it follows that:

$$n^{1/2}\hat{\Delta}_T^\mu \xrightarrow{d} \mathcal{N}\left(\Delta_T^\mu, V_{\Delta_T}\right)$$

and

$$V_{\Delta_T} = \frac{1}{\rho}V_0 + \frac{1}{1-\rho}V_1.$$

Analogously, to compute the asymptotic distribution of $\hat{\Delta}_S = \left(\hat{\beta}_1 - \hat{\beta}_0\right)'\hat{\mu}_X^{(1)}$ we know that:

$$n^{1/2}\begin{pmatrix} \left(\hat{\beta}_1 - \hat{\beta}_0\right) - \left(\beta_1 - \beta_0\right) \\ \hat{\mu}_X^{(1)} - \mu_X^{(1)} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} \frac{\boldsymbol{\sigma}_\beta}{\rho(1-\rho)} & . \\ \frac{\boldsymbol{\sigma}_{\beta\mu_X}^{(1)}}{(1-\rho)} & \frac{\boldsymbol{\sigma}_{\mu_X}^{(1)}}{(1-\rho)} \end{pmatrix}\right)$$

where $\boldsymbol{\sigma}_\beta = \rho\boldsymbol{\sigma}_\beta^{(1)} + (1-\rho)\,\boldsymbol{\sigma}_\beta^{(0)}$.

Now, define $\Delta_\beta = \beta_1 - \beta_0$. For the structure effect we have:

$$n^{1/2}\left(\hat{\Delta}_S^\mu - \Delta_S^\mu\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\Delta_S}\right)$$

and

$$V_{\Delta_S} = \frac{1}{1-\rho}\Delta_\beta'\boldsymbol{\sigma}_{\mu_X}^{(1)}\Delta_\beta + \frac{2}{1-\rho}\Delta_\beta'\boldsymbol{\sigma}_{\beta\mu_X}^{(1)}\mu_X^{(1)} + \frac{1}{\rho\,(1-\rho)}\mu_X^{(1)\prime}\boldsymbol{\sigma}_\beta\mu_X^{(1)}.$$

In a similar way, for the composition effect we get:

$$n^{1/2}\left(\hat{\Delta}_C^\mu - \Delta_C^\mu\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\Delta_C}\right)$$

and

$$V_{\Delta_C} = \frac{1}{\rho}\Delta'_{\mu_X}\boldsymbol{\sigma}_\beta^{(0)}\Delta_{\mu_X} + \frac{2}{\rho}\beta_0'\boldsymbol{\sigma}_{\beta\mu_X}^{(0)}\Delta_{\mu_X} + \frac{1}{\rho\left(1-\rho\right)}\beta_0'\boldsymbol{\sigma}_{\mu_X}\beta_0$$

with $\Delta_{\mu_X} = \mu_X^{(1)} - \mu_X^{(0)}$ and $\boldsymbol{\sigma}_{\mu_X} = \rho\boldsymbol{\sigma}_{\mu_X}^{(1)} + \left(1-\rho\right)\boldsymbol{\sigma}_{\mu_X}^{(0)}$.

Additionally, t-statistics can be constructed by plugging-in the empirical analogs of the corresponding variances. For instance:

$$t_{\Delta_T} = \frac{\hat{\Delta}_T^\mu}{\sqrt{\hat{\mathbb{V}}\Delta_T^\mu}} \xrightarrow{d} \mathcal{N}\left(0,1\right)$$

with $\hat{\mathbb{V}}\Delta_T^\mu = \frac{1}{n}\hat{V}_{\Delta_T} = \frac{1}{n_0}\hat{V}_0 + \frac{1}{n_1}\hat{V}_1$.

Validity of the Counterfactual Operator based on Cox model

The validity of the estimation and inference procedure of the CCOX follows the arguments in Chernozhukov et. al. (2013, CFM, hereafter). Consider the following regularity condition (which corresponds to Condition D in CFM):

Assumption 1.6 *Let $\mathcal{F}$ be a class of bounded measurable functions under the metric $\phi^{(\ell)}$ defined as:*

$$\phi^{(\ell)} = \left[\int\left(f-\tilde{f}\right)^2 dF^{(\ell)}\left(x\right)\right]^2$$

The following regularities hold:

a. Define the empirical processes:

$$\hat{M}^{(\ell)}(t,x) = n_\ell^{1/2}\left(\hat{F}^{(\ell)}(t|x) - F^{(\ell)}(t|x)\right)$$

$$\hat{N}^{(\ell)}(f) = n_\ell^{1/2}\int f d\left(\hat{F}^{(\ell)}(x) - F^{(\ell)}(x)\right)$$

with $f \in \mathcal{F}$. Then:

$$\left(\hat{M}^{(\ell)}(t,x), \hat{N}^{(\ell)}(f)\right) \Longrightarrow \left(M^{(\ell)}(t,x), N^{(\ell)}(f)\right)$$

where $\left(M^{(\ell)}(t,x), N^{(\ell)}(f)\right)$ is a zero mean tight Gaussian process, $M^{(\ell)}$ has uniformly continuous paths with respect to a standard metric on $\mathbb{R}^{1+k}$ and $N^{(\ell)}$ has uniformly continuous paths with respect to the metric $\phi^{(\ell)}$ on $\mathcal{F}$.

b. The map $t \mapsto F^{(\ell)}(t|.)$ is uniformly continuous with respect to the metric $\phi^{(\ell)}$.

To establish validity of the estimation and inference procedure based on bootstrapping methods, it is needed to verified the fulfillment of two high-level requirements, namely: *i.* the estimator of both conditional distribution and covariates distribution converge at parametric rate and satisfy a functional central limit theorem; and *ii.* bootstrapping methods are valid for estimating the limit laws of the conditional and the covariates distributions.

Under requirement *i.*, the counterfactual operator satisfies a functional central limit theorem, while requirements *i.* and *ii.* jointly guarantee that bootstrapping techniques are valid for making inference of the counterfactual operator and its smooth related functionals. The latter result is pertinent since the limit process of the counterfactual operator is nonpivotal.

Condition 1.6 is verified by Tsiatis (1981) and Andersen and Gill (1982). In particular, Tsiatis (1981) shows that $n_\ell^{1/2}\left(\hat{\beta}_\ell - \beta_\ell\right)$ converges in distribution to a normal random variable with zero mean, while the random function $n_\ell^{1/2}\left(\hat{\Lambda}_0^{(\ell)}(t) - \Lambda_0^{(\ell)}(t)\right)$ converges weakly to a Gaussian process (Theorems 3.2 and 6.1, respectively). These asymptotic results have also been documented by Naes (1982); Bailey (1983, 1984); Gill (1984). Following Tsiatis (1981, Lemma 6.2), we have that:

$$n_\ell^{1/2}\left(\hat{\Lambda}_0^{(\ell)}(t)\exp\left(\hat{\beta}_\ell' x\right) - \Lambda_0^{(\ell)}(t)\exp\left(\beta_\ell' x\right)\right) \Longrightarrow \mathcal{V}_x^{(\ell)}(t)$$

$$n_\ell^{1/2}\left(1 - \exp - \left(\hat{\Lambda}_0^{(\ell)}(t)\exp\left(\hat{\beta}_\ell' x\right)\right) - F^{(\ell)}(t|x)\right) \Longrightarrow \mathcal{S}_x^{(\ell)}(t)$$

where $\mathcal{S}_x(t)$ is a Gaussian process with zero mean and covariance structure given by:

$$\mathbb{C}ov\left(\mathcal{S}_x^{(\ell)}(t), \mathcal{S}_x^{(\ell)}(z)\right) = F^{(\ell)}(t|x)\,F^{(\ell)}(z|x)\,\mathbb{C}ov\left(\mathcal{V}_x^{(\ell)}(t), \mathcal{V}_x^{(\ell)}(z)\right) \quad 0 \leq t \leq z \leq \tau_H^{(\ell)}$$

Consequently, CCOX estimator satisfies a functional central limit theorem (that follows from CFM, Theorem 4.1). In addition, since $F^{(\ell)}(t|x)$ is Hadamard differentiable with respect to β_ℓ and $\Lambda_0^{(\ell)}(.)$ (see for details Freitag and Munk, 2005; McLain and Ghosh, 2011; Chen et. al., 2010; Hirose, 2011), and hence, by the chain rule of Hadamard differentiable maps (Van der Vaart and Wellner, 2004, Lemma 3.9.3), the counterfactual operator is Hadamard differentiable respect its arguments. Hence, the related smooth functionals also obey a central limit theorem (see Corollary 4.2 in CFM for details).

With respect to the inferential procedure, Cheng and Huang (2010) justify the validity of exchangeable resampling methods for general semiparametric $M-$estimators, which includes the Cox model as particular case. This verifies the second high-level requirement. As Corollaries 5.3 and 5.4 in CFM state, this shows that bootstrap consistently estimates the limit laws of the counterfactual operator based on the Cox model. Using the aforementioned argument, by Hadamard differentiability, this result holds for their smooth functionals.

1.7.2 Tables

Table 1.6 Performance OB Decomposition

Δ_C				
Censoring Level	**Sample Size**	**OB**	**OB - censored**	**COB**
	50	0.141	0.141	0.141
0%	500	0.045	0.045	0.045
	2500	0.020	0.020	0.020
	50	0.184	0.167	0.168
30%	500	0.151	0.082	0.056
	2500	0.149	0.075	0.026

Δ_S				
Censoring Level	**Sample Size**	**OB**	**OB - censored**	**COB**
	50	0.201	0.201	0.201
0%	500	0.064	0.064	0.064
	2500	0.027	0.027	0.027
	50	0.228	0.226	0.241
30%	500	0.155	0.098	0.080
	2500	0.148	0.074	0.036

Note: Numbers correspond to the average absolute deviations respect to the theoretical values of the composition and structure effect. OB: Oaxaca-Blinder decomposition. OB-censored: Oaxaca-Blinder decomposition dropping censored observations. COB: Censored Oaxaca-Blinder decomposition.

Table 1.7 Comparison between Kaplan-Meier and CCOX Estimators

$T \perp C$							
Censoring Level	**Sample Size**	**Kaplan-Meier**			**CCOX**		
		MD	**AD**	**MSE**	**MD**	**AD**	**MSE**
0%	**50**	0.00	0.00	0.00	26.81	9.20	0.39
	500	0.00	0.00	0.00	7.61	2.29	0.11
	2500	0.00	0.00	0.00	2.39	0.62	0.03
5%	**50**	23.54	5.54	0.32	34.92	11.75	0.50
	500	7.82	1.81	0.09	10.36	3.00	0.13
	2500	2.52	0.53	0.03	3.23	0.83	0.04
20%	**50**	94.91	22.99	1.13	97.53	24.43	1.16
	500	40.91	7.62	0.39	41.05	7.69	0.39
	2500	17.62	2.69	0.15	17.72	2.69	0.14
50%	**50**	226.17	54.30	2.62	226.79	53.66	2.61
	500	147.41	24.28	1.34	147.06	23.52	1.32
	2500	101.99	12.53	0.79	101.98	12.21	0.78

$T \perp C \mid X$							
Censoring Level	**Sample Size**	**Kaplan-Meier**			**CCOX**		
		MD	**AD**	**MSE**	**MD**	**AD**	**MSE**
0%	**50**	0.00	0.00	0.00	26.77	9.20	0.39
	500	0.00	0.00	0.00	7.62	2.29	0.11
	2500	0.00	0.00	0.00	2.39	0.62	0.03
5%	**50**	24.75	7.10	0.37	32.71	10.69	0.46
	500	14.06	4.52	0.20	9.72	2.64	0.12
	2500	11.21	3.80	0.17	3.04	0.72	0.03
20%	**50**	80.70	29.71	1.30	54.24	15.98	0.70
	500	57.23	21.80	0.92	15.67	4.01	0.18
	2500	51.56	18.81	0.84	4.93	1.14	0.05
50%	**50**	212.58	82.98	3.50	114.75	32.68	1.48
	500	162.56	65.64	2.78	36.54	8.68	0.39
	2500	150.83	57.97	2.56	12.86	2.72	0.13

Note: MD: Maximum Distance. AD: Absolute Deviation. MSE: Mean Squared Error. Numbers are multiplied by 1000 to facilitate comparisons.

Table 1.8 Performance CCOX Estimator: $T \perp C$

Weibull Times							
Censoring Level	**Sample Size**	**MD**		**AD**		**MSE**	
		B	**KP**	**B**	**KP**	**B**	**KP**
0%	**50**	26.75	22.52	9.18	8.22	0.63	0.60
	500	7.61	7.48	2.30	2.23	0.32	0.32
	2500	2.39	2.38	0.62	0.61	0.17	0.17
5%	**50**	34.84	31.29	11.71	10.51	0.71	0.68
	500	10.34	10.18	3.00	2.91	0.36	0.36
	2500	3.22	3.21	0.83	0.82	0.20	0.19
20%	**50**	98.03	93.86	24.46	23.39	1.08	1.05
	500	41.18	40.45	7.69	7.59	0.62	0.62
	2500	17.73	17.46	2.68	2.66	0.38	0.38
50%	**50**	227.41	225.05	53.73	53.09	1.62	1.61
	500	147.01	146.76	23.51	23.51	1.15	1.15
	2500	101.68	101.59	12.19	12.23	0.89	0.89

Normal Times							
Censoring Level	**Sample Size**	**MD**		**AD**		**MSE**	
		B	**KP**	**B**	**KP**	**B**	**KP**
0%	**50**	14.62	7.46	4.99	2.59	0.22	0.13
	500	3.87	4.28	1.40	1.30	0.06	0.06
	2500	3.85	3.91	1.08	1.06	0.05	0.05
5%	**50**	28.25	23.46	6.97	5.40	0.36	0.29
	500	8.61	8.18	2.01	1.99	0.10	0.09
	2500	4.17	4.20	1.17	1.15	0.05	0.05
20%	**50**	82.44	78.41	18.64	17.79	0.96	0.90
	500	27.45	26.66	5.07	4.92	0.27	0.26
	2500	9.74	9.50	1.84	1.78	0.10	0.09
50%	**50**	162.23	157.60	36.83	36.07	1.83	1.78
	500	62.71	61.33	10.87	10.63	0.59	0.57
	2500	23.91	23.18	3.66	3.53	0.21	0.20

Note: MD: Maximum Distance. AD: Absolute Deviation. MSE: Mean Squared Error. B: Breslow estimator. KP: Kalbfleisch-Prentice estimator. Numbers are multiplied by 1000 to facilitate comparisons.

Table 1.9 Performance CCOX Estimator: $T \perp C|X$

Weibull Times							
Censoring Level	**Sample Size**	**MD**		**AD**		**MSE**	
		B	**KP**	**B**	**KP**	**B**	**KP**
0%	**50**	26.79	22.55	9.20	8.22	0.39	0.36
	500	7.61	7.48	2.29	2.22	0.11	0.10
	2500	2.39	2.39	0.62	0.61	0.03	0.03
5%	**50**	32.76	29.80	10.69	9.55	0.46	0.41
	500	9.70	9.54	2.63	2.54	0.12	0.12
	2500	3.04	3.04	0.72	0.72	0.03	0.03
20%	**50**	54.32	51.78	16.02	14.85	0.70	0.65
	500	15.68	15.43	4.01	3.92	0.18	0.17
	2500	4.94	4.93	1.14	1.13	0.05	0.05
50%	**50**	115.32	111.99	32.76	31.71	1.49	1.43
	500	36.56	36.08	8.69	8.61	0.39	0.39
	2500	12.87	12.75	2.72	2.71	0.13	0.13

Normal Times							
Censoring Level	**Sample Size**	**MD**		**AD**		**MSE**	
		B	**KP**	**B**	**KP**	**B**	**KP**
0%	**50**	14.61	7.46	4.98	2.59	0.22	0.13
	500	3.88	4.28	1.40	1.31	0.06	0.06
	2500	3.85	3.91	1.08	1.06	0.05	0.05
5%	**50**	26.56	21.89	6.78	5.11	0.34	0.27
	500	7.72	7.52	1.87	1.88	0.09	0.09
	2500	4.17	4.23	1.09	1.08	0.05	0.05
20%	**50**	75.22	72.44	17.62	16.96	0.88	0.84
	500	24.07	23.83	4.70	4.65	0.24	0.24
	2500	8.30	8.30	1.65	1.64	0.08	0.08
50%	**50**	149.78	146.90	35.71	35.27	1.74	1.70
	500	52.93	52.52	9.88	9.82	0.51	0.51
	2500	18.81	18.75	3.20	3.19	0.17	0.17

Note: MD: Maximum Distance. AD: Absolute Deviation. MSE: Mean Squared Error. B: Breslow estimator. KP: Kalbfleisch-Prentice estimator. Numbers are multiplied by 1000 to facilitate comparisons.

Table 1.10 Decomposition Exercise: Mean Lifetime and Quartiles

Confidence Level	Censoring Levels		Truncated Mean		Q(0.50)	
	$\Pr\left(\delta^{(0)}=0\right)$	$\Pr\left(\delta^{(1)}=0\right)$	**Percentile**	**Hybrid**	**Percentile**	**Hybrid**
95%	**0%**	**0%**	0.961	0.962	0.958	0.953
	0%	**30%**	0.954	0.963	0.952	0.940
	30%	**0%**	0.963	0.972	0.958	0.944
	30%	**30%**	0.952	0.966	0.968	0.944
90%	**0%**	**0%**	0.907	0.913	0.917	0.911
	0%	**30%**	0.902	0.911	0.915	0.903
	30%	**0%**	0.915	0.923	0.915	0.897
	30%	**30%**	0.912	0.917	0.907	0.895

Confidence Level	Censoring Levels		Q(0.25)		Q(0.75)	
	$\Pr\left(\delta^{(0)}=0\right)$	$\Pr\left(\delta^{(1)}=0\right)$	**Percentile**	**Hybrid**	**Percentile**	**Hybrid**
95%	**0%**	**0%**	0.946	0.928	0.957	0.935
	0%	**30%**	0.965	0.945	0.958	0.940
	30%	**0%**	0.968	0.942	0.964	0.931
	30%	**30%**	0.963	0.933	0.958	0.930
90%	**0%**	**0%**	0.907	0.882	0.909	0.869
	0%	**30%**	0.926	0.897	0.920	0.896
	30%	**0%**	0.925	0.897	0.920	0.884
	30%	**30%**	0.916	0.886	0.909	0.884

Table 1.11 Dependence of Censoring on Covariates: Goodness-of-fit Probability Models

	Exit from Unemp.		Unemp. to Emp.	
	Women	**Men**	**Women**	**Men**
Linear Prob. Model	0.046	0.083	0.177	0.123
Logit	0.070	0.128	0.166	0.167

Note: Authors' calculations.

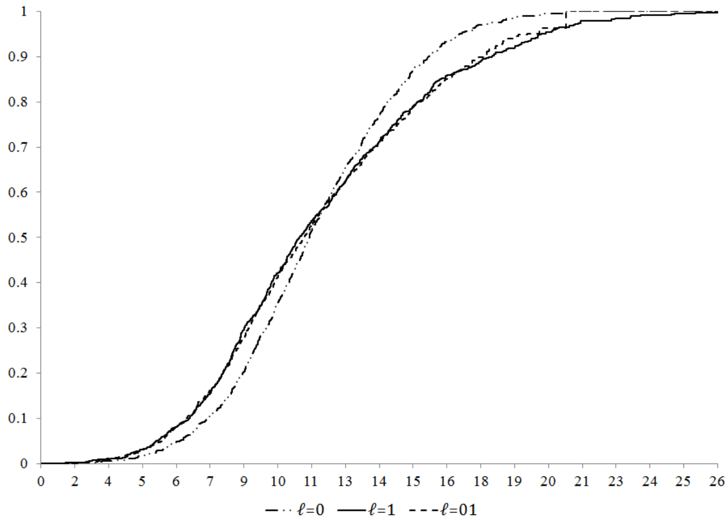
Table 1.12 Decomposition of Mean Difference Ignoring Censoring

	Exit from Unemp.		**Unemp. to Emp.**	
	COB	**CCOX**	**COB**	**CCOX**
Total	2.085	2.040	0.945	0.876
Composition	0.438	0.471	0.089	0.029
Structure	1.647	1.569	0.857	0.847

Note: Authors' calculations.

1.7.3 Figures

Figure 1.1 Decomposition Exercise: Simulated Data



2 Distributional Regression in Survival Analysis

2.1 Introduction

Existing survival models usually assume a parametric or semiparametric conditional hazard function (CHF) depending on parameters easily interpreted and typically estimated by maximum likelihood methods; e.g. the proportional hazard model (Cox, 1972, 1975), the propotional odds model (Clayton, 1976; Bennett, 1983; Murphy et. al., 1997), or the accelarated failure model (Kalbfleisch and Prentice, 1980). See Guo and Zeng (2014) for a overview on semiparametric models in survival analysis. In these models, the CHF indentifies the conditional distribution function (CDF) expressed in terms of a known error distribution of the transformed survival time linear model.

This chapter proposes inference procedures for the CDF of survival times controlling for covariates, under a distributional regression (DR) specification using randomly right censored data. The DR models appeared as an alternative to the Koenker and Basset (1978) quantile regression (QR) model to characterize the CDF. Foresi and Peracchi (1995) introduced a DR model as a natural extension of the logistic regression to characterize the CDF by estimating the coefficients at each distribution point using binary choice regression models. Discussion about the relative merits of QR and DR is formely presented in Perachi (2002); Koenker et. al. (2013) and Chernozhukov et. al. (2013).

Both QR and DR models became popular tools in econometrics to perform counterfactual analysis, mainly for the estimation of policy effects (see e.g. Bitler et. al., 2006; Angrist et. al., 2006; Rothe, 2010; Frandsen et. al., 2012; Frölich and Melly, 2013; Donald and Hsu, 2014) and counterfactual decompositions (see e.g. Machado and Mata, 2005; Fortin et. al., 2011; Rothe, 2012; Chernozhukov et.

al., 2013). However, these methods are not directly implementable in presence of random censoring, a common feature in survival data.

Consequently, other relevant target variables related to durations, like unemployment duration, have not been considered yet. Exceptions are Garcia-Suaza (2016) and Sant'Anna (2016), who considered counterfactual analysis survival time with an underlying purely nonparametric CDF. Inference methods based on QR models using survival data under random censoring have been proposed by Honore et. al. (2002); Portnoy (2003); Peng and Huang (2008); Wang and Wang (2009); Gorfine et. al. (2014), among others. QR methods involve to deal with nonsmooth and nonconvex objective functions and require to implement iterative estimation procedures that might be computational demanding. Nonetheless, no attention has been paid to the DR model in survival analysis, though it seems particularly well motivated.

We propose to estimate the DR coefficients by a plug-in method, where the infeasible empirical joint distribution of survival times and covariates is substituted by the corresponding Kaplan-Meier type estimator, a method proposed by Stute (1993) for estimating parameters of multivariate distributions when one of them is subject to right censoring. That is, DR coefficients estimator is obtained by weighted binary choice regression using Kaplan-Meier weights, which are computationally as easy to compute as the corresponding estimates in the absence of censoring. We provide a central limit theorem for the coefficient estimators under fairly weak regularity conditions. Unlike QR estimators, the asymptotic variance of DR estimators can be computed without resorting to smooth estimates of the conditional density of survival times.

We find that the DR modelling strategy nests the most popular CHF specifications, by making the coefficients to depend on the underlying duration. Therefore, this provides robust alternatives for estimating the CDF. It also provides an efective way of checking the specification of popular survival models by testing that some DR coefficients are constant. Moreover, one of the main advantages of our proposal is its computational tractability: it is fully data-driven, does not require choosing tuning parameters such as bandwidths, can accommodate both discrete and continuous data, and avoids the curse of dimensionality. Lastly, the estimation procedure can be easily adapted using conventional statistical softwares.

Next section introduces the basic notation and motivates DR survival models from regression models of transformed survival times. Third section presents the estimation method. Fourth section discusses sufficient regularity conditions for justifying inferences on the estimated conditional distribution. Fifth section reports the results of a small Monte Carlo experiment, where we compare the relative performance of DR with classical survival models such as the proportional hazard and the proportional odds. We also apply the proposed techniques to two studies of unemployment duration using US data. We investigate the causal effect of unemployment benefits on duration and perform a counterfactual decomposition comparing pre-crisis and post-crisis in the context of the Great Depression. Last section suggests some extensions for further work. Proofs are in the mathematical Appendix.

2.2 Survival Distributional Regression

Consider the $\mathbb{R}^{2+k}-valued$ random vector (T, C, X) defined on $(\Omega, \mathcal{F}, \mathbb{P})$, where T is survival time, C is a censoring variable and X is a vector of covariates. In the context of survival analysis, T might be observed under censoring due to lack of follow-up. In particular, collecting duration data requires following individuals over time and it might occur that individuals either do not change their status during the follow-up period or withdraw.

However, data is available on (Y, X, δ), where $Y = \min(T, C)$ and $\delta = 1_{\{T \leq C\}}$, with $1_{\{.\}}$ the indicator function. Survival analysis methods have focused in modelling hazard rates, and then, the underlying distribution function. A crucial function to this end is the cumulative conditional hazard rate, which is given by

$$\Lambda_{T|X}(t|X) := \int_0^t \frac{F_{T|X}(dt|X)}{1 - F_{T|X}(t - |X)},$$

where $F_{T|X}$ is the conditional distribution of T given X, and for any generic function J, $J(t-) = \lim_{s \uparrow t} J(t)$. Assuming that $F_{T|X}$ is continuous in t, which is sensible in continuous time survival analysis, $\Lambda_{T|X}(t|X) = -\ln\left(1 - F_{T|X}(t|X)\right)$ a.s., and also

$$\frac{F_{T|X}(dt|X)}{1 - F_{T|X}(t - |X)} = \frac{F_{T|X}(t + dt|X) - F_{T|X}(t|X)}{1 - F_{T|X}(t|X)} = P(T < t + dt|X, T \geq t) \ \ a.s. \tag{2.1}$$

This is the conditional probability, given X, than an individual who has survived (or is at risk) at time t will not be alive after $t + dt$. If T is unemployment duration, equation (2.1) is the probability that an unemployed individual at period t with characteristics X will find a job before $t + dt$. If also $F_{T|X}$ is differentiable in t with Lebesgue density $f_{T|X}$,

$$P(T < t + dt|X, T \geq t) = \frac{f_{T|X}(t|X)\,dt}{1 - F_{T|X}(t|X)} =: \lambda_{T|X}(t|X)\,dt \text{ as } dt \to 0 \ a.s.,$$

where $\lambda_{T|X}$ is the conditional hazard rate. This is the instantaneous failure rate controling for covariates, i.e. in the unemployment duration context, $\lambda_{T|X}(t|x)$ is the probability that an unemployed person at t with characteristics x will find a job inmediately.

The conditional distribution of any random variable can be identified by means of a linear transform regression model (see Horthon, Kneib and Bühlmann 2014 for discussion). Consider the regression model

$$\varrho(T) = X'\beta_0 + F_0^{-1}(U) \ a.s., \tag{2.2}$$

where "$'$" means transpose, β_0 is a vector of unknown parameters, ϱ is a known monotonically increasing transformation function, F_0 is a distribution function, and for any generic function J, $J_0^{-1}(u) = \inf\{u : J_0(u) \geq u\}$, i.e. F_0^{-1} is the quantile function of F_0, and U is independent of X. We may consider an intercept term, though it typically cannot be identified in these models. The discussion is centered in the linear model for simplicity, but a nonlinear regression model could also be considered. Therefore, the CDF, and hence, the CHF is identified when F_0 is known. The corresponding CDF is

$$F_{T|X}(t|X) = F_0(\varrho(t) - X'\beta_0). \tag{2.3}$$

Notice that, for a fixed t, this is a binary regression model of $1_{\{T\leq t\}}$ on X, i.e. $\mathbb{E}\left(1_{\{T\leq t\}}|X\right) = F_{T|X}(t|X)$ $a.s.$. See Doksum and Gasko (1990) for discussion on the correspondence between models in binary regression and in survival analysis.

Many models in survival analysis can be characterized by a linear transformation model like equation (2.2) and with conditional distribution as (2.3). For instance, the Cox (1972, 1975) proportional hazard (PH) model specifies $\lambda_{T|X}(t|X) = \lambda_0(t)\exp(-X'\beta_0)$, where λ_0 is a nonparametric baseline function. This corresponds to model (2.2) with $\varrho(t) = \log\Lambda_0(t)$ and $\Lambda_0(t) = \int_0^t \lambda_0(v)dv = -\ln(1 - F_0(t))$. In turn, the acelerated failure time model proposed by Kalbfleisch and Prentice (1980) specifies $\lambda_{T|X}(t|X) = \lambda_0(t\exp(-X'\beta_0))\exp(-X'\beta_0)$, where $\lambda_0(.)$ is the hazard rate function of $\exp\left(F_0^{-1}(U)\right)$, which corresponds to model (2.2) with $\varrho(t) = \log t$ and F_0 is usually assumed a log-logistic distribution. When F_0 is a Weibull distribution, the hazard function is monotone and if F_0 is an extreme error distribution then λ_0 is a constant, i.e., the hazard rate of the exponential distribution. The proportional odds (PO) model (Clayton, 1976; Bennett, 1983; Murphy et. al., 1997) specifies the conditional odds on death function $\Gamma_{T\,X|}\left(t|X\right) = F_{T|X}\left(t|X\right)\left[1 - F_{T|X}\left(t|X\right)\right]^{-1}$, which is of the form $\Gamma_{T\,X|}\left(t|X\right) = \Gamma_0(t)\exp\left(-X'\beta_0\right)$, where $\Gamma_0(u) = F_0(u)/\left[1 - F_0(u)\right]$ is the baseline odds on death function. This corresponds to (2.2) with $\varrho\left(t\right) = \log\Gamma_0(t)$. When F_0 the logistic distribution, i.e. $F_0(u) = \left[1 + e^{-u}\right]^{-1}$ and $\Gamma_0(u) = e^u$ the PH and PO models are identical.

Taking $\varrho\left(t\right) = t$ in (2.2) and F_0 unknown, the $\tau - th$ conditional quantile of T given X is

$$F_{T|X}^{-1}\left(\tau|X\right) = X'\beta_0 + F_0^{-1}(\tau),\ \tau \in \left(0,1\right),$$

where F_0 is nonparametric. This is a QR model with identical slope coefficients β_0 and intercept $F_0^{-1}(\tau)$ at each quantile $\tau \in \left(0,1\right)$. Koenker and Basset (1978) generalize this model by allowing varying slope coefficients β_0, which is the case when the errors in the regression model are heteroskedastic. The resulting QR model is,

$$F_{T|X}^{-1}\left(\tau|X\right) = \mathbf{X}'\alpha(\tau),\ \tau \in \left(0,1\right), \tag{2.4}$$

where $\mathbf{X} = (1, x')'$ and the vector of functions $\alpha(\tau)' := \left(F_0^{-1}(\tau), \beta_0\left(\tau\right)\right)$ is nonparametric. The main econometrics interest is the vector of slope functions $\beta_0\left(\tau\right)$. An interesting feature this model is that whole quantile function can be characterized by estimating separate models for a grid of τ's (Melly, 2005) or combining all quantiles in one joint model (He, 1997; Schnabel and Eilers, 2013).

Model (2.3) has serious limitations. For instance, in the context of modelling unemployment duration, a relevant explanatory variable is the amount and duration of unemployment benefits. Let us consider $X = (X_1, X_2')'$, where X_1 is unemployment benefits amount and X_2 the rest of explanatory variables. Suppose that unemployment benefits are exhausted after t_0 periods. Then, a reasonable model explayning the effect of unemployment benefits on unemployment duration is given by Cox (1972, 1975) with covariates involving time dependent components, with corresponding conditional cumulative hazard

$$\Lambda_{T|X}\left(t|\, X\right) = \Lambda_0(t) \exp\left(-X_1 1_{\{t \leq t_0\}} \beta_{01} - X_2' \beta_{02}\right), \tag{2.5}$$

with a time-dependent variable $X_1 1_{\{t \leq t_0\}}$. Other interactive effects between $1_{\{t \leq t_0\}}$, X_1 and X_2 may also be considered. For instance, we could consider the extra covariate $X_1 t^{-1}$ to indicate that unemployment benefits have effect even after exhaustion. Model (2.5), with time varying coefficients provides a way of explaning the causal effect of unemployment benefit along the unemployment history without considering time dependent explanatory variables.

$$\Lambda_{T|X}\left(t|\, X\right) = \Lambda_0(t) \exp\left(-X_1 \beta_{01}(t) - X_2' \beta_{02}\right),$$

with $\beta_{01}(t) = \varpi_{01} 1_{\{t \leq t_0\}} + \varpi_{02} t^{-1}$, and $\varpi_0' s$ are parameters. However, it is sensible to leave $\beta_{01}(.)$ unspecified, i.e. nonparametric interactive effects. In fact, the function $\beta_{10}\left(.\right)$ provides valuable information on the evolution of the causal effects of unemployment benefits during unemployment duration. This motivates introducing time-varying coefficient models in survival analysis as Foresi and Peracchi (1995) DR models.

We can specify a conditional odds-to-death function

$$\Gamma_{T|X}\left(t|\, X\right) = \Psi_0\left(\mathbf{X}'\theta(t)\right) a.s.,$$

with $\mathbf{X} = (1, X')'$ and $\theta(t) = (\varrho(t), \beta(t)')'$ for some given function $\Psi_0 : \mathbb{R} \to \mathbb{R}$. If, $\Psi_0(z) = e^z$,

$$F_{T|X}(t|X) = \frac{\exp\{\mathbf{X}'\theta(t)\}}{1 + \exp\{\mathbf{X}'\theta(t)\}}, \tag{2.6}$$

which corresponds to the propotional odds model with $\beta(t) = \beta_0$ $a.s.$ We can also specify $F_{T|X}$ directly, i.e.

$$F_{T|X}(t|X) = F_0(\mathbf{X}'\theta(t)), \tag{2.7}$$

for a logistic $F_0(t) = [1 + e^{-t}]^{-1}$, which corresponds to (2.6). Notice that we can check the specification of any classical model in the direction (2.7), by testing that the $X's$ coefficents are constant.

Since, for t fixed equation (2.7) is a binary regression model, identification of $\theta(t)$ for each $t \in \mathbb{Z}$, with $\mathbb{Z}$ denoting the support of T, is achieved in terms of a loss function measuring the loss between $1_{\{T \leq t\}}$ and the probability $F_{T|X}(t|X)$ under the DR specification. That is, for each $t \in \mathbb{Z}$, $1_{\{T \leq t\}}$ is distributed, conditionally on X, as a Bernoulli random variable with parameter $F_0(\mathbf{X}'\theta(t))$ $a.s.$,

$$\theta(t) = \arg\min_{\theta \in \mathbb{R}^{1+k}} \mathbb{E}\left[\rho\left(1_{\{T \leq t\}}, F_0(\mathbf{X}'\theta)\right)\right], \tag{2.8}$$

where $\rho(u, v) = -u \ln v - (1 - u)\ln(1 - v)$ is the Kullback-Leibler loss function. Other loss functions, like the square loss $\rho(u, v) = (u - v)^2$, or the absolute error loss $\rho(u, v) = |u - v|$, could be used. Note that smothness of the conditional density is not required since the approximation is done pointwise. Natural estimators are obtained by pluging-in a consistent estimator of the joint distribution of (T, X) into the expectations in equation (2.8). But, the sample distibution of (T, X) cannot be used because of the censoring. Instead, we propose to plug-in the Kaplan-Meier estimator studied by Stute (1993, 1996).

2.3 Distributional Regression Model under Censoring

In the presence of right random censoring we observe a random sample $\{Y_i, X_i, \delta_i\}_{i=1}^n$ of (Y, X, δ). A natural estimator of

$$\theta(t) = \underset{\theta\in\mathbb{R}^{1+k}}{\arg\min} \int_{\mathbb{R}^{1+k}} \rho\left(1_{\{\bar{t}\leq t\}}, F_0\left((1,\bar{x})'\theta\right)\right) F_{T,X}(d\bar{t}, d\bar{x}), \tag{2.9}$$

where $F_{T,X}(t,x) := \mathbb{P}(T \leq t, X \leq x)$ is the joint distribution of (T, X), consists of plugging-in some consistent estimator of $F_{T,X}$ obtained from the observed sample. The sample distribution of (T, X) would be an obvious candidate if $\{T_i, X_i\}_{i=1}^n$ were observed.

Of course, we can only consistently estimate $F_{T,X}$ when it is identified. Independence between C and T is necessary for consistently estimating the marginal distribution of T, which forms a basis for the Kaplan-Meier estimator. Stute (1993) founded that the extension to the multivariate case, including covariates, needs an extra restriccion of the dependence of C and X, as stated below.

Assumption 2.1 *The following conditions hold:*
a. $\mathbb{P}(T \leq t, C \leq c) = \mathbb{P}(T \leq t)\,\mathbb{P}(C \leq c)$.
b. $\mathbb{P}(T \leq C|T, X) = \mathbb{P}(T \leq C|T)$.

This assumption has been widely used in survival analysis (c.f. Stute, 1999; Uña-Álvarez and Rodríguez-Campos, 2004; Sanchez-Sellero et. al., 2005; Sant'Anna, 2016; Garcia-Suaza, 2016). Assumption 2.1.a. is the classical independence assumption that guarantees identification of the marginal distribution of survival times (c.f. Peterson, 1977). In turn, Assumption 2.1.b. states the relation between the censoring mechanism and the covariates so that, given the actual survival times

T, the covariates do not provide any further information on whether censoring occurs (see Stute and Wang, 1993 for further discussion). In this framework, potential dependence between C and X is allowed, and of course, it is also held when C is independent of (T, X).

Next proposition states the main result on identification of $F_{T,X}$ in terms of the subdistributions of (Y, X, δ), $H_d(t, x) = \mathbb{P}(Y \leq t, X \leq x, \delta = d)$, which can be estimated from available data. To this end, we introduce the function

$$\Lambda(t, x) = \int_0^t \frac{F_{T,X}(d\bar{t}, x)}{1 - F_{T,X}(\bar{t}-, \infty)}, \quad (t, x) \in \mathbb{R}^{1+k}, \tag{2.10}$$

where we use the notation $\infty = (\infty, ..., \infty)'$. From the multiplicative relation between the survival function and the agregate hazard function,

$$1 - F_{T,X}(t, x) = \exp\left(-\Lambda^c(t, x)\right) \prod_{\bar{t} \leq t} \left[1 - \Lambda\left(\{\bar{t}\}, x\right)\right], \tag{2.11}$$

where $\Lambda^c(., x)$ is the continuous part of $\Lambda(., x)$, and for any generic fucntion J, $J(\{t\}) = J(t) - J(t-)$. Define $H(t) := H_1(t, \infty) + H_0(t, \infty) = \mathbb{P}(Y \leq t)$.

Proposition 2.1 *Under Asumption 2.1,*

$$\Lambda(t, x) = \int_0^t \frac{H_1(d\bar{t}, x)}{1 - H(\bar{t}-)}.$$

$H_d(t, x)$ can be estimated by its sampling distribution

$$\hat{H}_d(t, x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{Y_i \leq t,\ X_i \leq x,\ \delta_i = d\}},$$

and $\hat{H}(t) = \hat{H}_1(t,\infty) + \hat{H}_0(t,\infty)$. The corresponding estimator of Λ is

$$\hat{\Lambda}(t,x) = \int_0^t \frac{\hat{H}_1(d\bar{t},x)}{1-\hat{H}(\bar{t}-)} = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{Y_i\le t, X_i\le x, \delta_i=1\}}}{1-\hat{H}(Y_i-)} = \sum_{i=1}^{n} \frac{1_{\{Y_i\le t,\delta_i=1\}}}{n-R_i+1} 1_{\{X_i\le x\}},$$

where R_i is the rank of Y_i among $Y_1, ..., Y_n$.

According to equation (2.11), the survival function corresponding to $\hat{\Lambda}$ is

$$1-\hat{F}_{T,X}(t,x) = \prod_{\bar{t}\le t}\left[1-\hat{\Lambda}\left(\{\bar{t}\},x\right)\right],$$

which is the extension of the Kaplan and Meier (1958) estimator. The jump size equals $\hat{\Lambda}(\{Y_i\},x) = [n-R_i+1]^{-1} 1_{\{X_i\le x\}}$, and hence,

$$1-\hat{F}_{T,X}(t,x) = \prod_{Y_i\le t, X_i\le x, \delta_i=1} \frac{n-R_i}{n-R_i+1} = \prod_{Y_i\le t, X_i\le x}\left(1-\frac{\delta_i}{n-R_i+1}\right).$$

After ordering the $Y_i' s$, we obtain,

$$1-\hat{F}_{T,X}(t,x) = \prod_{Y_{i:n}\le t, X_{[i:n]}\le x}\left(1-\frac{\delta_{[i:n]}}{n-i+1}\right),$$

where $Y_{1:n} \le ... \le Y_{n:n}$ and for generic random vectors $\{\xi_i\}_{i=1}^n$, $\xi_{[i:n]}$ is the i-th ξ-concomitant of $Y_{i:n}$, i.e. $\xi_{[i:n]} = \xi_j$ if $Y_{i:n} = Y_j$. Therefore, the mass given to $(Y_{i:n}, x)$ is, for each $x \in \mathbb{R}^k$,

$$W_i = \left[\hat{F}_{T,X}(Y_{i:n},x) - \hat{F}_{T,X}(Y_{i-1:n},x)\right] = \frac{\delta_{[i:n]}}{n-i+1}\left[1-\prod_{j=1}^{i-1}\left(1-\frac{\delta_{[j:n]}}{n-i+1}\right)\right].$$

In case of ties, it is considered as if uncensored observations precede the censored observations, and other kind of ties are ordered arbitrarily. In the absence of censoring, and $\{T_i, X_i\}_{i=1}^n$ are observed, $W_i = n^{-1}$ and $\hat{F}_{T,X}$ is the sample distribution.

So, we can express $\hat{F}_{T,X}$ in the additive form (see Corollary 1 in Garcia-Suaza, 2016 for details),

$$\hat{F}_{T,X}(t,x) = \sum_{i=1}^{n} W_i 1_{\left\{Y_{i:n} \leq t, X_{[i:n]} \leq x\right\}}.$$

The main purpose of $\hat{F}_{T,X}$ consists of estimating functionals of the form $\mathbb{E}\left[\varphi(T,X)\right] = \int_{\mathbb{R}^{1+k}} \varphi(t,x) F_{T,X}\left(dt, dx\right)$, for a given integrable function $\varphi : \mathbb{R}^{1+k} \to \mathbb{R}$, by the Kaplan-Meier integral,

$$\int_{\mathbb{R}^{1+k}} \varphi(t,x) \hat{F}_{T,X}\left(dt, dx\right) = \sum_{i=1}^{n} W_i \varphi(Y_{i:n}, X_{[i:n]}),$$

which is a weighted average with weights W_i. Stute (1993, 1996) provide under Assumption 2.1 and certain restrictions on φ, consistency and a central limit theorem for this kind of integrals (called Kaplan-Meier integrals). Hence, the estimator of $\theta(t)$ is obtained replacing $\hat{F}_{T,X}$ in equation (2.9), and the resulting estimator is,

$$\begin{aligned}
\hat{\theta}(t) &= \underset{\theta \in \mathbb{R}^{1+k}}{\arg\min} \int_{\mathbb{R}^{1+k}} \rho\left(1_{\{\bar{t} \leq t\}}, F_0\left(\bar{\mathbf{x}}'\theta\right)\right) \hat{F}_{T,X}(d\bar{t}, d\bar{x}) \\
&= \underset{\theta \in \mathbb{R}^{1+k}}{\arg\min} \sum_{i=1}^{n} W_i \rho\left(1_{\{Y_{i:n} \leq t\}}, F_0\left(\left(1, X_{[i:n]}\right)' \theta\right)\right).
\end{aligned}$$

Then the corresponding CDF estimator is,

$$\hat{F}_{T|X}\left(t|x\right) = F_0\left(\mathbf{X}'\hat{\theta}(t)\right). \tag{2.12}$$

Alternative semiparametric estimators of the conditional distribution have been developed under Assumption 2.1, but they ussually require smoothing and the choice of tunning parameters (see c.f. Bouaziz and Lopez, 2010 and Strzalkowska-Kominiak and Cao, 2013). Interestingly, in absence of covariates, $\hat{F}_{T|X}\left(t|x\right)$ reduces to the Kaplan-Meier estimator.

2.4 Asymptotic Theory and Inference

We focus on the Kullback-Leibler loss function, which corresponds to

$$\hat{\theta}(t) = \underset{\theta\in\Theta}{\arg\min} \quad \hat{Q}(t,\theta)$$

where

$$\hat{Q}(t,\theta) := \int_{\mathbb{R}^{1+k}} q(y,x;t,\theta)\hat{F}_{T,X}(dy,dx) = \sum_{i=1}^{n} W_i q_t(Y_{i:n}, X_{[i:n]};\theta),$$

estimates

$$Q\left(t,\theta\right) := \mathbb{E}\left[q(T,X;t,\theta)\right],$$

with

$$\begin{aligned} q(y,x;t,\theta) &= \rho\left(1_{\{y\leq t\}}, F_0\left((1,x)'\,\theta\right)\right) \\ &= -\left\{1_{\{y\leq t\}} \ln F_0\left((1,x)'\,\theta\right) - 1_{\{y>t\}} \ln\left[1-F_0\left((1,x)'\,\theta\right)\right]\right\}. \end{aligned}$$

Though the results presented bellow can also be applied for other loss functions. First, we provide $\sqrt{n}-$consistency of $\theta(t)$ at each $t\in\mathbb{Z}$. Regularity conditions are those needed for the consistency of parameters in binary regression (e.g. Amemiya, 1985, Section 9.2.2). Define,

$$\dot{q}(T,X;t,\theta) = \frac{F_0\left(\left(\mathbf{X}'\theta\left(t\right)\right)\right) - 1_{\{T\leq t\}}}{F_0\left(\mathbf{X}'\theta\left(t\right)\right)\left[1-F_0\left(\mathbf{X}'\theta\left(t\right)\right)\right]} f_0\left(\mathbf{X}'\theta\left(t\right)\right)\mathbf{X}.$$

and

$$A\left(\theta\right) = \mathbb{E}\left[\frac{f_0^2\left(\mathbf{X}'\theta(t)\right)}{F_0\left(\mathbf{X}'\theta(t)\right)\left[1-F_0\left(\mathbf{X}'\theta(t)\right)\right]}\mathbf{X}\mathbf{X}'\right].$$

Assumption 2.2 *The following conditions hold:*

a. $F_{Y|X}(t|X) = F_0(\mathbf{X}'\theta(t))$ *a.s. for all* $t \in \mathbb{Z}$ *and some function* $\theta : \mathbb{R}^{+} \rightarrow \mathbb{R}$.

b. F_0 *is twice continuously differentiable, with derivative* f_0 *and second order derivative* f_0', *for all* $u \in \mathbb{R}$ *and* $f_0(\mathbf{X}'\theta) > 0$, $F_0(\mathbf{X}'\theta) \in (0,1)$ *a.s.*

c. For each $t \in \mathbb{Z}$, $\theta(t)$ *is an interior point of* Θ, *a compact subset of* $\mathbb{R}^{1+k}$ *.*

d. $A(\theta)$ *is p.d. for every* $\theta \in \Theta$.

Conditions 2.2.a. and 2.2.b. are satisfied for F_0 logistic (logit), standard normal (probit) or extreme value distribution. Conditions 2.2.c. and 2.2.d. guarantee existence and identifiability of $\hat{\theta}(t)$. In most cases, $\hat{\theta}(t)$ is the solution of the equation,

$$\frac{\partial}{\partial\theta}\hat{Q}(t,\theta) = \int_{\mathbb{R}^{1+k}} \dot{q}(\bar{t},\bar{x};t,\theta)\hat{F}_{T,X}(d\bar{t},d\bar{x}) = \sum_{i=1}^{n} W_i\dot{q}(Y_{i:n}, X_{[i:n]};t,\theta) = 0$$

Consider $\tau_H = \inf\{t : H(t) = 1\}$, which is the least upper bound for the support of the marginal distribution of Y. For most lifetime distributions considered in the literature, $\tau_H = \infty$. Next Proposition applies Stute (1993, 1999) results to obtain the convergence of $\hat{Q}_t(\theta)$ uniformly in $\theta \in \Theta$, which is essential to prove consistency. Then, Theorem in Stute (1993) and result for uniform convergence of integrals, see e.g. Jennrich (1969) Theorem 3, establish that,

Proposition 2.2 *Under Assumption 2.2.a-c, for each* $t \leq \tau_H$, *and for any* $t \in \mathbb{Z}$ *when the support of* $F(.,\infty)$ *is unbounded,*

$$\sup_{\theta\in\Theta}\left|\left(\hat{Q} - Q\right)(t,\theta)\right| = o(1) \ a.s.$$

Notice that, by Proposition 2.2 and dominated convergence, uniformly in $\theta \in$

Θ,

$$\frac{\partial}{\partial\theta}\hat{Q}(t,\theta) = \frac{\partial}{\partial\theta}Q\left(t,\theta\right) + o(1)\ a.s$$

where

$$\frac{\partial}{\partial\theta}Q\left(t,\theta\right) = \mathbb{E}\left[\dot{q}(T,X;t,\theta)\right] = \mathbb{E}\left[\frac{F_0\left(\mathbf{X}'\theta\right) - F_0\left(\mathbf{X}'\theta\right)}{F_0\left(\mathbf{X}'\theta\right)\left[1 - F_0\left(\mathbf{X}'\theta\right)\right]}\right] \text{ for all } t \leq \tau_H,$$

which vanishes when $\theta = \theta(t)$. Furthermore, since $\partial^2 Q(t,\theta)/\partial\theta\partial\theta' = A(\theta)$ is p.d. for all $\theta \in \Theta$ by Assumption 2.2.d, $\theta(t)$ is the only point in the parameter space Θ that minimizes $Q(t,\theta)$. This justifies the consistency of $\hat{\theta}(t)$.

Asymptotic normality is proved under Conditions 2.2.a-c. Next Proposition, based on Stute (1996, 1999), provides an asymptotic expansion of $\partial\, \hat{Q}(t,\theta(t)) \big/ \partial\theta$, which forms a basis to derive the asymtotic distribution of $\hat{\theta}(t)$. Define

$$\psi\left(Y,X,\delta;t,\theta\right) := \dot{q}\left(Y,X;t,\theta\right)\gamma_0(Y)\delta + \gamma_1(Y;\theta)(1-\delta) - \gamma_2\left(Y;\theta\right),$$

with

$$\begin{aligned}
\gamma_0\left(y\right) &= \exp\left\{\int_0^{y-} \frac{H_0\left(d\bar{t},\infty\right)}{1 - H(\bar{t})}\right\}, \\
\gamma_1\left(y;\theta\right) &= \frac{1}{1 - H(y)}\int 1_{\{y<\bar{t}\}}\varphi\left(\bar{t},\bar{x};\theta\right)\gamma_0\left(t\right)dH_1\left(d\bar{t},d\bar{x}\right), \\
\gamma_2\left(y;\theta\right) &= \int \frac{1_{\{\bar{t}<y,\bar{t}<\tilde{y}\}}\varphi(\bar{y},\bar{x})\gamma_0\left(\tilde{y}\right)}{\left[1 - H(\tilde{y})\right]^2} H_0\left(d\tilde{y},\infty\right)H_1\left(d\bar{t},d\bar{x}\right).
\end{aligned}$$

In order to justify the expansion, we need two extra assumptions. Define,

$$K\left(y\right) = \int_0^{y-} \frac{G\left(d\bar{t}\right)}{\left[1 - H(\bar{t})\right]\left[1 - G\left(\bar{t}\right)\right]}.$$

Assumption 2.3 *Suppose the following conditions hold,*

a. $\mathbb{E}\left|\psi\left(Y,X,\delta;t,\theta\right)\gamma_0\left(Y\right)\delta\right|^2 < \infty$ *for each* $t \in \mathbb{Z}$.

b. $\mathbb{E}\left|\psi\left(Y,X,\delta;t,\theta\right)\sqrt{K(Y)}\right|^{2}<\infty$ *for each* $t\in\mathbb{Z}$.

Proposition 2.3 *Under Assumptions 2.1-2.3, if* $t\leq\tau_{H}$, *or for all* $t\in\mathbb{Z}$ *if* $F(.,\infty)$ *has unbounded support,*

$$\frac{1}{n}\sum_{i=1}^{n}\psi\left(Y_{i},X_{i},\delta_{i};t,\theta(t)\right)+\left(\hat{\theta}-\theta\right)(t)'A(\theta)=o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

The asymptotic distribution of $\hat{\theta}(t)$ is an inmediate consequence of Proposition 2.3. Define $B(t_i,t_j):=\mathbb{E}\left[\psi\left(Y,X,\delta;t_i,\theta(t_i)\right)\psi\left(Y,X,\delta;t_j,\theta(t_j)\right)\right]$ and $V(t_i,t_j)=A\left(\theta(t_i)\right)^{-1}B(t_i,t_j)A\left(\theta(t_j)\right)^{-1}$, $t_i,t_j\in\mathbb{R}^{+}$.

Corollary 2.1 *Under Assumptions 2.1-2.3,*

$$\sqrt{n}\left(\left(\hat{\theta}-\theta\right)(t_1),\ldots,\left(\hat{\theta}-\theta\right)(t_m)\right)\xrightarrow{d}\left(Z(t_1),\ldots,Z(t_m)\right),$$

for $(t_1,\ldots,t_m)\in\mathbb{R}^m$, *where* $Z(t_1),\ldots,Z(t_m)$ *are* $k+1-$*valued random vectors with mean zero and covariance matrix*

$$V(t_i,t_j)=\mathbb{E}\left[Z(t_i)Z(t_j)'\right]=A\left(\theta(t_i)\right)^{-1}B(t_i,t_j)A\left(\theta(t_j)\right)^{-1}\quad i,j=1,\ldots,m.$$

The covariance matrix $V(t_i,t_j)$ can be estimated without resorting to smooth estimates of nonparametric components. In particular,

$$\hat{V}(t_i,t_j)=\hat{A}\left(\theta(t_i)\right)^{-1}\hat{B}(t_i,t_j)\hat{A}\left(\theta(t_j)\right)^{-1}\quad i,j=1,\ldots,m.$$

where,

$$\hat{A}\left(\theta\right)=\frac{1}{n}\sum_{i=1}^{n}\frac{f_0^2\left(\mathbf{X}_i'\theta(t)\right)}{F_0\left(\mathbf{X}_i'\theta(t)\right)\left[1-F_0\left(\mathbf{X}_i'\theta(t)\right)\right]}\mathbf{X}_i\mathbf{X}_i'$$

and

$$\hat{B}(t_i,t_j) := \frac{1}{n}\sum_{i=1}^{n} \psi\left(Y,X,\delta;t_i,\theta(t_i)\right)\psi\left(Y,X,\delta;t_j,\theta(t_j)\right).$$

Therefore the basis of the inference procedure consists on computing the sample analog of ψ, which depend on γ_0, γ_1 and γ_2. Moreover, once we compute the conditional distribution, one might be interest in functionals with complex structures. In such case, estimating variances could become cumbersome so that inference based on nonparametric (empirical) bootstrapping techniques is an appealing alternative.

Finally, in order to test a classical specification, it suffices to test that $\beta(t)$ is constant, where $\theta(t) = \left(\varrho(t), \beta(t)'\right)'$ i.e., $F_{T|X}(t|X) = F_0\left(\varrho(t) + \beta(t)'\mathbf{X}\right)$. Corollary 2.1 offers a way of testing hypothesis at a number of point. That is, testing

$$H_0 : \beta(t_i) = \beta(t_j), \quad i,j = 1,\ldots,m$$

$$H_0 : \beta(t_i) \neq \beta(t_j), \text{ some } i,j = 1,\ldots,m.$$

The test statistic is

$$J = n \begin{pmatrix} \hat{\beta}(t_1) - \hat{\beta}(t_m) \\ \vdots \\ \hat{\beta}(t_{m-1}) - \hat{\beta}(t_m) \end{pmatrix}' \hat{\mathbb{V}}^{-1} \begin{pmatrix} \hat{\beta}(t_1) - \hat{\beta}(t_m) \\ \vdots \\ \hat{\beta}(t_{m-1}) - \hat{\beta}(t_m) \end{pmatrix}$$

where $\hat{\mathbb{V}}$ has components estimating the asymptotic variances of $\sqrt{n}\left(\hat{\beta}(t_j) - \hat{\beta}(t_m)\right)$, $j = 1,\ldots,m$, i.e. $AsyVar\left(\sqrt{n}\left(\hat{\beta}(t_j) - \hat{\beta}(t_m)\right)\right)$ is estimated by $\hat{V}(t_j,t_j) - \hat{V}(t_j,t_m) - \hat{V}(t_m,t_j) + \hat{V}(t_m,t_m)$.

2.5 A Monte Carlo Study

In order to evaluate the finite sample properties of the DR estimator, we compare its performance with respect to classical survival regression models, such as the PH and PO. We consider three designs: $DGP^{(1)}$ a PH, $DGP^{(2)}$ a PO and $DGP^{(3)}$ a DR model where hazard depends on time-varying coefficients. We conduct a Monte Carlo experiment for 1000 replications with sample size 100, 500 and 1500, and censoring levels of 0% and 30%. The three designs are described below.

For $DGP^{(1)}$ and $DGP^{(2)}$,

$$F_{T|X}\left(t|\,X\right) = F_0(\varrho(t) - \beta_0 X),$$

with X distributed $U(0,1)$ and $\beta_0 = 1$. $DGP^{(1)}$ is a PH model with $F_0(u) = 1 - \exp(-(u/\eta)^{\alpha})$ with $\alpha = 1$ and $\eta = 2$, i.e. a Weibull(1,2), $\varrho(t) = \log \Lambda_0(t)$, and the censoring variable C distributed is Weibull(1,1.8) to generate the 30% censoring. The corresponding CHF is $\lambda_{T|X}\left(t|\,X\right) = 2t\exp\left(2\beta_1 X\right)$, which is monotone in t. $DGP^{(2)}$ is a PO with $F_0(u) = \left[1 + (u/\alpha)^{\eta})\right]^{-1}$ with $\alpha = 1$ and $\eta = 4$, i.e. a Log-logistic(1,4), $\varrho(t) = \log \Gamma_0(t)$ with $\Gamma_0(u) = F_0(u)/\left[1 - F_0(u)\right]$, with C distributed Log-logistic(0.22,0.3) in order to obtain a 30% censoring. The corresponding CHF is $\lambda_{T|X}\left(t|\,X\right) = 4t^3/\left[t^4 + \exp\left(\beta_2 X\right)\right]$.

For $DGP^{(3)}$, we consider

$$F_{T|X}\left(t|\,X\right) = F_0\left(\theta(t)X\right),$$

with $\theta(t) = (t^2/2 - 1)$ and $F_0(u) = 1 - \exp(-\exp\left(u\right))$, i.e. an extreme value distribution, X distributed $U(0.1, 1.1)$ and C distributed Weibull(3,2) in order to obtain 30% censoring. The corresponding CHF is $\lambda_{T|X}\left(t|\,X\right) = \exp\left(\ln t + \ln X + \theta(t)X\right)$.

We study the robustness of CDF estimates based on CR specification which are compared with estimates of classical survival models with constant slope coefficients. First, we consider Cox (1972, 1975) model, which corresponds to (2.3) with $\varphi(t) = \log \Lambda_0(t)$. This is equivalent to assume that F_0 belongs to a extreme value distribution family, i.e.

$$F_{T|X}(t|X) = 1 - \exp(-\exp(\log \Lambda_0(t) - X\beta)).$$

This PH specification is estimated by plugging-in the partial likelihood estimator of β and the Breslow (1974) estimator of $\Lambda_0(t)$. Second, we consider a PO, which corresponds to (2.3) with $\varphi(t) = \log \Gamma_0(t)$ and F_0 logistic, i.e.

$$F_{T|X}(t|X) = \frac{1}{1 + \exp(-\log \Gamma_0(t) - X\beta)}$$

where $\Gamma_0(t)$ and β are estimated using Hunter and Lange (2002) method, which considers $\Gamma_0(t)$ unrestricted but monotone.

The DR model is estimated assuming that F_0 in (2.7) is logistic, i.e.,

$$F_{T|X}(t|X) = \frac{1}{1 + \exp(-\theta_1(t) - \theta_2(t)X)}$$

which is consistent for $DGP^{(2)}$ and $DGP^{(3)}$, and F_0 extrem value, i.e.,

$$F_{T|X}(t|X) = 1 - \exp(-\exp(\theta_1(t) - \theta_2(t)X))$$

which is consistent with $DGP^{(1)}$ and $DGP^{(3)}$.

The performance of each estimator of $F_{T|X}(t|X)$, $\tilde{F}_{T|X}(t|X)$ say, is examinated using the following measures

$$MD\left(x\right)=\max_{t}\left|\tilde{F}_{T|X}\left(t|x\right)-F_{T|X}\left(t|x\right)\right|,\ \ AD\left(x\right)=\frac{1}{n}\sum_{i=1}^{n}\left|\tilde{F}_{T|X}\left(t|x\right)-F_{T|X}\left(t|x\right)\right|.$$

That is, the maximum distance (MD) and the average absolute distance (AD). Comparisons are made based on the average over the 1000 draws. Lastly, note that DR estimates are not necessarily monotone in t, so that some improvement can be achieved through monotone rearrangement procedures. To do so, we use the algorithms proposed by Chernozhukov et. al. (2013) and Donald and Hsu (2014). Denote $\tilde{F}^{CFG}_{T|X}\left(t|x\right)$ and $\tilde{F}^{DH}_{T|X}\left(t|x\right)$ the rearranged version of the conditional distribution for a given $x\in\mathbb{R}$ for the Chernozhukov et. al. (2013) and Donald and Hsu (2014) procedures, respectively. Thus,

$$\tilde{F}^{CFG}_{T|X}\left(t|x\right)=\inf\left\{u:\int 1_{\left\{\tilde{F}_{T|X}(s|x)\leq u\right\}}ds\geq t\right\},$$

and, define $\tilde{F}^{DH}_{T|X}\left(t|x\right)=\tilde{F}_{T|X}\left(t|x\right)$ for $0\leq t<Y_{1:n}$ and $\tilde{F}^{DH}_{T|X}\left(Y_{n:n}|x\right)=1$ and for any other $Y_{i:n}\leq t<Y_{i+1:n}$,

$$\begin{aligned}\tilde{F}^{DH}_{T|X}\left(t|x\right) &= 1_{\left\{\tilde{F}_{T|X}(Y_{i:n}|x)>1\right\}}+1_{\left\{0\leq\tilde{F}_{T|X}(Y_{i:n}|x)\leq\tilde{F}^{DH}_{T|X}(Y_{i-1:n}|x)\right\}}\tilde{F}^{DH}_{Y|X}\left(Y_{i-1:n}|x\right)\\ &\quad +1_{\left\{\tilde{F}^{DH}_{T|X}(Y_{i-1:n}|x)<\tilde{F}_{T|X}(Y_{i:n}|x)\leq 1\right\}}\tilde{F}_{Y|X}\left(Y_{i:n}|x\right).\end{aligned}$$

We first compare the aforementioned methods at the median of X, $x_{med}=0.5$. Table 2.1 presents the average of the AD (multiplied by 1000) across 1000 draws. The DR estimator is computed using a grid of 200 points applying the DH rearrangement. In general, performance of all estimators improves as sample size increases; however, important differences are observed among the DGPs. Under $DGP^{(1)}$, PH estimation model outperforms but the difference respect to the DR estimator is rather than slight and vanishes with the sample size. Same result can be observed for $DGP^{(2)}$. Remarkably the estimation under PH specification exhibits very similar results to the estimation under PO.

Table 2.1 Average Distance (DH) Classical Survival Models and DR

	Cens. level	Sample size	PH	PO	$\mathbf{F_0}$(Logistic)	$\mathbf{F_0}$(Ext.value)
$DGP^{(1)}$	**0%**	**100**	21.98	23.84	25.12	24.58
		500	7.79	13.02	10.08	9.04
		1500	3.99	10.30	6.15	5.16
	30%	**100**	33.23	44.31	42.15	41.47
		500	12.56	16.89	16.91	15.66
		1500	6.65	13.40	10.26	8.48
$DGP^{(2)}$	**0%**	**100**	21.79	22.09	23.97	23.96
		500	6.74	6.75	7.63	7.71
		1500	3.04	3.02	4.21	4.38
	30%	**100**	33.23	37.86	44.99	45.09
		500	13.93	13.99	18.17	18.12
		1500	7.56	7.66	10.44	10.49
$DGP^{(3)}$	**0%**	**100**	34.42	43.17	23.65	25.28
		500	30.43	40.33	12.29	15.44
		1500	28.90	38.40	10.37	14.05
	30%	**100**	41.23	58.00	33.76	35.11
		500	33.67	42.68	15.82	18.97
		1500	31.25	40.40	12.50	16.49

Note: Conditional distribution is evaluated at $x = 0.5$. DR estimates are rearranged using DH algorithm. AD multiplied by 1000 to facilitate comparisons.

For $DGP^{(3)}$, DR estimator outperforms the other two alternative, which are not consistent, for the censored and no-censored cases. In fact, no important gain of increasing sample size is observed for the PH and PO methods. Moreover, results are roughly the same for the two chosen baseline distributions. Similar results are obtained if we compare the re-arrangement algorithms and the distance based on MD (see Tables 2.4 and 2.5).

To have a broader evaluation, the same execise was run for a grid of 100 point of the covariate's support. ADs for 1500 observations are plotted in Figure 2.2. For both cases, censoring and no-censoring, results are qualitatively the same. Noticeably, under classical assumptions ($DGP^{(1)}$ and $DGP^{(2)}$), all models estimate fairly similar. As it is natural, PH model outperforms when proportionality assumption holds, but there is little differences with respect to the DR, although using the extreme value link function turns out in more accurate estimates. The same is observed for $DGP^{(2)}$, where the PO specification and the DR with logistic link function present relatively lowest ADs. An interesting result is that DR estimator always works better that the alternative specification.

In contrast, PH and PO models reduce significantly their accuracy under $DGP^{(3)}$. In fact, the difference with respect to the DR estimator is even remarkable for some observed outliers, which mainly corresponds to extreme values of the covariate distribution. This gives strong evidence in favor of the robustness of the DR estimator to capture complex functional forms.

2.6 Studying Unemployment Duration in the US

To show the flexibility of the DR estimator to analyze different features of survival times distribution, we implement this method to make inference on duration to exit to unemployment in the US using two different microdata sets. First, the effect of unemployment benefits on duration for liquidity constrained and unconstrained households is examined following the empirical exercises in Chetty (2008). Secondly, we investigate the changes in unemployment duration during the recent so-called Great Recession. That is, by estimating the conditional distribution before and after the crisis, we analyze the heterogeneity of the effect of relevant demographic variables and perform a decomposition à la Oaxaca-Blinder to quantify the contribution of observed socioeconomic characteristics on the cyclical dynamics of unemployment outflows.

2.6.1 Unemployment Benefits, Liquidity Constraints and Unemployment duration

One of the main concerns in the design of unemployment benefits (UB) policies is the adverse effect generated on unemployment duration. The channel explaining this link is that receiving unemployment benefits induces moral hazard by distorting the relative prices of leisure and consumption and reducing the incentives to search job. In this context, Chetty (2008) studies the liquidity constraints as an additional channel that enlarges unemployment duration for UB receivers. That is, UB increases cash-in-hand and consumption for those workers unable to smooth consumption, and reduces the pressure to find a job, implying that labor supply is more sensitive to transitory income shocks in presence of imperfections in credit and insurance markets.

In order to test this conjecture, we use the DR estimator to compute the effect of UB on unemployment duration for whole unemployed population, and for different groups defined according to potential differences in liquidity constraints. We compare our results with the empirical analysis of Chetty (2008), who estimates Cox models including an interaction between the parameter of interest and the underlying duration, which turns out no significant. One interesting advantage of the DR is that non-parametric behavior of the coefficient can be captured. In particular, we estimate

$$F_{T|X}\left(t|\,UB,\mathbf{Z}\right)=F_0\left(\theta_1(t)+\theta_2(t)\log UB+\theta_3(t)\mathbf{Z}\right)$$

where UB is the received amount of unemployment benefits, $\mathbf{Z}$ is a set of individual characteristics and F_0 is logistic. We are particularly interested in $\theta_2(t)$.

To do so, it is considered data from the Survey of Income and Program Participation (SIPP) for the period 1985-2000. SIPP collects information on household and individual characteristics, as well as employment status and UB receipt. Sample consists on prime-age males that suffer a job separation and report to be job seekers. Since the ability to smooth consumption can not be observed, assets holdings and an indicator of having to make a mortgage payment are used as proxy of liquidity constraints. See Chetty (2008, section 3) for further details.

Therefore, we estimate the survival distribution of unemployment duration using the DR. In addition to the average weekly unemployment benefit (in logs), we control by age, marital status, education and total household wealth. This model is estimated for the total sample and by group according to net wealth (by tertiles) and the presence of mortgage payments. We first study the behavior of the parameter associated to UB over duration spell (see Figure 2.3 in Appendix). For all sample, it is observed that the effect of UB on survival probability is positive and roughly constant over unemployment spell, which agrees with findings of Chetty (2008). More interesting results are obtained by comparing *the most* liquidity constrained and *the less* liquidity constrained unemployed. For the first tertile, the coefficient is always significant and higher compared with whole sample, while in the highest tertile, it turns out mostly no significant.

Interestingly, in contrast to Chetty (2008), UB seems to have a temporal effect on unemployment duration for middle wealth workers. In particular, the coefficient is significant up to the probability of being unemployed more than 16 weeks and exhibits a decreasing trend over unemployment spell. In turn, when liquidity constraints are defined according mortgage payments, same general result is obtained since effect of UB on unemployment duration is significant for workers experiencing liquidity constraints. Overall, this empirically supported the differential impact of UB benefit for workers unable to smooth consumption.

One appealing advantage of implementing the DR estimator is that the effect of UB on whole survival distribution and average unemployment duration (and even for other features of the conditional distribution) are feasible to compute. To do so, we define a synthetic worker by setting all controls at the average and the UB at 160 USD weekly. Thus, we compute the change in the survival distribution and average unemployment duration of increasing UB by 20%.

Figure 2.4 show changes in survival distribution, including 90% bootstrap confidence intervals. As expected, results follow the same reasoning obtained in the coefficients. For all sample, it is estimated a average change in survival probability of 1.52 p.p. Furthermore, hump shape effect is found for the first two tertiles. The higher UB, the lower probability of leaving unemployment. This effect has a positive trend in the early weeks, but decreases later on. In the case of the first tertile, the marginal effect on survival probability increases for long time and exhibit a sort of hysteresis effect, since decreasing slowly over unemployment duration. In the second tertile, the effect vanishes rapidly. Similar result is observed when mortgage payments are present.

While Chetty (2008) finds a negative effect on survival probability, although no significant, for the less constrained unemployed, our estimation suggest that UB could have some significant effect for medium term unemployment. Based on the estimated conditional distribution, we compute the average unemployment duration. It is observed that the more liquidity constrained, the highest variation un average duration (see Table 2.2). While for all sample the estimated variation is 3.6%, for the most constrained workers the effect is rather than double (8.61% and 7.60% for less wealthy and mortgage payers, respectively).

Table 2.2 Effect of Unemployment Benefits on Unemployment Duration

Sample		Baseline	20%+ UB	Change Duration
All		20.698	21.445	3.61%
Tertile net wealth	**1st**	20.334	22.086	8.61%
	2nd	20.649	21.353	3.41%
	3rd	19.456	19.452	-0.02%
Mortgage	**No**	21.587	21.708	0.56%
	Yes	19.589	21.078	7.60%

Note: Authors' calculations. Baseline corresponds to the average duration for the average worker.

2.6.2 Unemployment Duration in the Great Recession

The strong consequences of the recent economic crisis on the labor market performance have aroused the interest of studying the determinants of labor market flows, mainly the unemployment exits. Recent studies have focused on analyzing the transition rate from unemployment through linear probability models, which usually require a parametrization of the underlying duration dependence. In this context, instead of the hazard function, we use the DR estimator to estimate the survival distribution of unemployment duration conditional on relevant demographic characteristics for periods before and after the economic crisis.

To do so, we use information from the Current Population Survey (CPS) for the period 2005 to 2011. The CPS is a rotating panel following individuals during two periods of four consecutive months, with eight months in between. This structure facilitates the construction of unemployment duration based on the reported employment status. However potential misclassification of employment status might occur due to high turnover and recall, which is corrected following Fujita (2011). In addition, CPS also collects rich information on socioeconomic characteristics of workers. In our analysis, data set is constructed using the available codes by Rothstein (2011).

Pre-crisis and post-crisis periods are defined according to the date when workers start their unemployment spell. In particular, pre-crisis refers to those workers starting unemployment between January 2005 until June 2007, while post-crisis corresponds to unemployment spells from January 2009 to Febrary 2011. Population consists on the set of unemployed workers between 25 and 69 years old. Additional filters correspond to availability of the covariates set, which includes gender, marital status, age (cuadratic spline), education and state. In this case, the reference categories are male, single, and bachelor or higher, repectively.

In order to compared our results with classical methods, Cox models are also estimated[12]. We compute marginal effects of the conditional survival distribution for gender, marital status and education for the average unemployed worker for the pre-crisis period. That is, all covariates are fixed at their average while the one of interest changes from 1 to 0. For education the comparison is based on the extreme cases, i.e. less than high school vs bachelor or higher.

The resulting marginal effects are plotted in Figure 2.5, including 90% confidence intervals computed by bootstrapping. First interesting result is that marginal effects computed based on Cox model follows very similar trend (due to the proportionality assumption), while DR produces non-monotonic behavior over duration spells, mainly in the case of marital status. Our results suggest that women and married workers have lower surviving probability over whole unemployment period, and less educated workers seems to experience persistently higher probabilities of being unemployed, although is not significant in the pre-crisis period. Comparing the pre and post crisis periods, the most important differences are observed for the coefficients of education, where the marginal effect becomes significant and

[12] Proportionality assumption is tested through the Schoenfeld residual test, obtaining fairly low p-values, 0.108 and 0.095 for pre and post crisis preiods. In the detailed analysis, marital status and age report the lowest p-values.

increasing, and gender that doubles for the same period.

As second exercise, we exploit the observed changes the coefficients to compute a counterfactual decomposition. In particular, we seek to quantify to what extent the changes in coefficients and in the workers composition explain the difference the distributional features unemployment duration. To be specific, the target is to estimate

$$\Delta_T^{\phi} = \phi\left(F_T^{(1,1)}\right) - \phi\left(F_T^{(1,0)}\right) + \phi\left(F_T^{(1,0)}\right) - \phi\left(F_T^{(0,0)}\right) = \Delta_C^{\phi} + \Delta_S^{\phi}$$

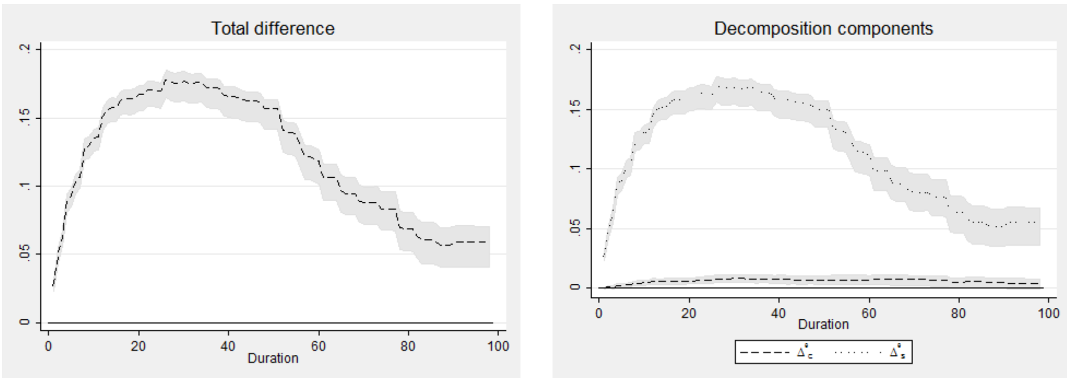
where ϕ is the functional of interest and $F_T^{(i,j)}$ is a counterfactual distribution given by $F_T^{(i,j)}(t) = \mathbb{E}\left[F(t|x, D=i)\,|D=j\right]$ with $D=1$ denoting the post-crisis period and $D=0$ the pre-crisis period. So, $F_T^{(i,j)}$ denotes de distribution of population j if they face the schedule of population i, i.e., $F_T^{(0,1)}$ is the unemployment duration distribution that we would observe if unemployed workers before crisis period would face circumstances observed after crisis.

By comparing functionals of these counterfactuals distributions, the total difference Δ_T^{ϕ} can be splitted into two components: the part explained by differences in the population composition Δ_C^{ϕ} (composition effect), and the part due to differences in the underlying conditional distribution Δ_S^{ϕ} (structure effect). Noticeably, ϕ encompasses real-valued parameters such as the median or well known inequality measures like the Gini coefficient, but also function-valued parameters such as the hazard function, the distribution function or the Lorenz Curve. For futher discussion on identification and estimation related to counterfactual decompositions see Fortin et. al. (2011), and for the case of durations outcomes Garcia-Suaza (2016).

Therefore, we use the DR estimator in order to compute the counterfactual distributions. That is, $\hat{F}_T^{(i,j)}(t) = \mathbb{E}\left[\hat{F}(t|x, D=i)\,|D=j\right]$. Similar decomposition exercise for the unemployment exit rate can be found in Bachmann and Sinning (2016). This approach allows us to make an appealing analysis since computing $\hat{F}_T^{(i,j)}$ enables to decompose the whole distribution as well as other functionals of interest like average unemployment duration or the probability of being long term unemployed.

Figure 2.1 shows the results for the distributional decomposition, that is $\phi = 1_{\{T \leq t\}}$. The total difference reveals that economic downturn generates higher persistence in unemployment. This effect is overall higher to 5 p.p. and it is heterogeneous over the unemployment period. Regarding the decomposition components, we observe that these follow similar shape than in the total, but with remarkable differences in magnitude. The two components are statistically significant, although composition effect is small. Bachmann and Sinning (2016) found same results for the transition rate.

Figure 2.1 Distributional Decomposition Unemployment Duration



Note: Authors' calculations. 90% confidence intervals computed by bootstrapping.

Based on the estimates of counterfactual distributions, we also compute the decomposition components for other functional of interest (see Table 2.3) such as the average unemployment duration, the Gini coefficient and the probability of being long term unemployed (LTU), i.e. the probability of being unemployed more that 26 or 52 weeks. As expected, total difference is positive for the average duration and the LTU, while in the case of the Gini coefficient is negative indicating that economics crisis tends to raise unemployment duration, mainly through a higher incidence of LTU. Composition and structure effects have the same sign and are significant. While structure effect represents direct impact of changes in the return of observed characteristics, composition effect suggests that changes in workers socioeconomic profile also contributes to raise unemployment duration. In other words, unemployment outflows are affected not only for the less favorable economic conditions, but also for the profile of the job seekers. Lastly, the composition effect represents around 5% of the total different for all studied functionals.

Table 2.3 Decomposition of Unemployment Duration Functionals

	Average	**LTU(26)**	**LTU(52)**	**Gini**
Pre-Crisis	37.542	0.424	0.280	0.541
Post-Crisis	49.310	0.602	0.421	0.433
Counterfactual	48.710	0.594	0.414	0.437
Δ^{ϕ}	11.768	0.178	0.141	-0.108
	[10.625 , 12.373]	[0.1649 , 0.1849]	[0.1247 , 0.1480]	[-0.111 , -0.100]
Δ^{ϕ}_{C}	0.601	0.008	0.007	-0.005
	[0.2801 , 0.8926]	[0.0048 , 0.0110]	[0.0029 , 0.0109]	[-0.006 , -0.002]
Δ^{ϕ}_{S}	11.168	0.170	0.134	-0.103
	[9.9652 , 11.818]	[0.1563 , 0.1779]	[0.1167 , 0.1410]	[-0.107 , -0.095]

Note: Authors' calculations. 90% confidence intervals computed by bootstrapping in [ ].

2.7 Final Remarks and Further Research

We have proposed a robust Distribution Regression procedure in the context of survival analysis, where the variable of interest is usually observed under censoring. Moreover, this model turns out to generalize the classical survival regression models that assume particular functional forms to the hazard function, to the case of time-varying coefficients.

In this way, resulting estimates of the conditional distribution are a useful tool to perform statistical inference either on the conditional distribution itself or on functionals of the underlying duration outcome. For instance, $F_{T|X}$ can be used to policy analysis in order to estimate average treatment effect and heterogeneous treatment effects (see c.f. Rothe, 2010). In this perspective, an interesting extension for the proposed procedure consists on regarding the case of endogenous covariates, as in the case of quantile regression models (see c.f. Chernozhukov and Hansen, 2005). Additionally, once conditional distribution is estimated, plugg-in estimators of quantile treatment effects, difference-in-difference estimator or regression discontinuity are possible to compute. Even though the model was motivated to the case of dependent duration variable, $\hat{F}_{T|X}$ also estimates the conditional distribution of uncensored outcomes depending on survival times.

Finally, an uniform CLT for $\sqrt{n}\left(\hat{\theta}-\theta\right)(t)$ is interesting to provide a specification classical test. Consider $\theta(t) = \left(\varrho(t), \beta(t)'\right)'$ i.e., $\mathbf{X}'\theta(t) = \varrho(t) + \beta(t)'\mathbf{X}$. An interesting specification test is

$$H_0 : \beta(t) = \bar{\beta},\ \forall t \in \mathbb{R}^k$$

where $\bar{\beta}$ is a vector of constants. So, once we get an uniform CLT for $\sqrt{n}\left(\hat{\theta}-\theta\right)(t)$ we can provide a test for H_0 in the direction of non-parametric alternatives based on the statistic

$$\hat{\eta} = n\left\|\hat{\beta}\right\|,$$

where $\|.\|$ is some suitable measure.

2.8 Appendix

2.8.1 Some Theoretical Results

Proof of Proposition 2.1

This proposition follows the same the arguments of Garcia-Suaza (2016). That is, under Assumption 2.1.b. we have:

$$\begin{aligned} H_1(t,x) &= \mathbb{P}(Y \le t, X \le x, \delta = 1) = \mathbb{P}(T \le t, X \le x, T \le C) \\ &= \int_0^t [1 - G(\bar{t}-)] F_{T,X}(d\bar{t}, x), \end{aligned}$$

and by Assumption 2.1.a.

$$1 - H(t) = \mathbb{P}(Y > t) = [1 - F_{T,X}(t-, \infty)][1 - G(t)].$$

Thus, using equation (2.10)

$$\Lambda(t,x) = \int_0^t \frac{F_{T,X}(d\bar{t}, x)}{1 - F_{T,X}(\bar{t}-, \infty)} = \int_0^t \frac{H_1(d\bar{t}, x)}{1 - H(\bar{t}-)}.$$

Proof of Proposition 2.2

The Theorem in Stute (1993) asserts that for each $t \in \mathbb{R}^+$,

$$\hat{Q}(t,\theta) = \mathbb{E}\left[q(Y,X,t,\theta) 1_{\{Y \le \tau_H\}}\right] + 1_{\{\tau_H \in M\}} \mathbb{E}\left[q(Y,X,t,\theta) 1_{\{Y > \tau_0\}}\right]$$

where M is the set of jumps of H, possibly empty. Then, since $F_0(\mathbf{X}'\theta(t))$ *a.s.*,

$$\mathbb{E}\left[\sup_{\theta \in \Theta} |q(Y,X,t,\theta)|\right] < \infty.$$

Then apply result on uniform convergence of integrals, e.g. Jennrich (1969) Theorem 2.

Proof of Proposition 2.3

Since $\theta(t)$ is an interior point of Θ, applying the mean value theorem, with probability one,

$$\begin{aligned}\frac{\partial}{\partial\theta_j}\hat{Q}\left(t,\hat{\theta}(t)\right) &= \sum_{i=1}^{n} W_i \dot{q}_j\left(Y_{i:n}, X_{[i:n]}, \Delta_{[i:n]}; t, \theta(t)\right) \\ &\quad + \sum_{i=1}^{n} W_i \left(\frac{f_0^2\left(\mathbf{X}_i'\tilde{\theta}^{(j)}\right)}{F_0\left(\mathbf{X}_i'\tilde{\theta}^{(j)}\right)\left[1-F_0\left(\mathbf{X}_i'\tilde{\theta}^{(j)}\right)\right]} + \tilde{\eta}_i^{(j)}(t) \right) X_{ij}\mathbf{X}_i'\left(\hat{\theta}-\theta\right)(t).\end{aligned}$$

for $\tilde{\theta}^{(j)} : \left\|\left(\tilde{\theta}^{(j)}-\theta\right)(t)\right\| \leq \left\|\left(\hat{\theta}-\theta\right)(t)\right\|$ $a.s.$, $j = 1, .., k+1$ and all $t \in \mathbb{R}^+$, where φ_j is the $j-th$ element of φ and

$$\tilde{\eta}_i^{(j)}(t) = \frac{\left[F_0\left(\mathbf{X}_i'\tilde{\theta}^{(j)}\right) - 1_{\{T\leq t\}}\right] f_0^2(\mathbf{X}_i'\tilde{\theta}^{(j)})}{F_0^2\left(\mathbf{X}_i'\tilde{\theta}^{(j)}\right)\left[1-F_0\left(\mathbf{X}_i'\tilde{\theta}^{(j)}\right)\right]^2}\left(2F_0\left(\mathbf{X}_i'\tilde{\theta}^{(j)}\right)-1\right)$$

Now,

$$\sum_{i=1}^{n} W_i \dot{q}_j\left(Y_{i:n}, X_{[i:n]}, \delta_{[i:n]}; t, \theta(t)\right) = \frac{1}{n}\sum_{i=1}^{n}\psi_j\left(Y_i, X_i, \delta_i; t, \theta(t)\right) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$$

$$\begin{aligned}\sum_{i=1}^{n} W_i \left(\frac{f_0^2\left(\mathbf{X}_i'\tilde{\theta}^{(j)}\right)}{F_0\left(\mathbf{X}_i'\tilde{\theta}^{(j)}\right)\left[1-F_0\left(\mathbf{X}_i'\tilde{\theta}^{(j)}\right)\right]} + \tilde{\eta}_i^{(j)} \right)\mathbf{X}_i X_{ij} &= \frac{1}{n}\sum_{i=1}^{n}\frac{f_0\left(\mathbf{X}_i'\theta(t)\right)}{F_0\left(\mathbf{X}_i'\theta(t)\right)\left[1-F_0\left(\mathbf{X}_i'\theta(t)\right)\right]}\mathbf{X}_i X_{ij} \\ &\quad + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)\end{aligned}$$

where ψ_j is the $j-th$ element of ψ, applying Stute (1996) Theorem 1.1.

2.8.2 Tables

Table 2.4 Average Distance (CFG) Classical Survival Models and DR

	Cens. level	Sample size	PH	PO	$\mathbf{F_0}$(Logistic)	$\mathbf{F_0}$(Ext.value)
$DGP^{(1)}$	**0%**	**100**	21.98	23.84	25.02	24.42
		500	7.79	13.02	10.07	9.02
		1500	3.99	10.30	6.15	5.16
	30%	**100**	33.23	44.31	41.97	41.20
		500	12.56	16.89	16.84	15.54
		1500	6.65	13.40	10.24	8.44
$DGP^{(2)}$	**0%**	**100**	21.79	22.09	23.94	23.92
		500	6.74	6.75	7.63	7.70
		1500	3.04	3.02	4.21	4.37
	30%	**100**	33.23	37.86	44.93	45.01
		500	13.93	13.99	18.15	18.09
		1500	7.56	7.66	10.43	10.47
$DGP^{(3)}$	**0%**	**100**	34.42	43.17	23.63	25.23
		500	30.43	40.33	12.30	15.44
		1500	28.90	38.40	10.38	14.05
	30%	**100**	41.23	58.00	33.62	34.95
		500	33.67	42.68	15.85	18.95
		1500	31.25	40.40	12.53	16.49

Note: Conditional distribution is evaluated at $x = 0.5$. DR estimates are rearranged using CFG algorithm. AD multiplied by 1000 to facilitate comparisons.

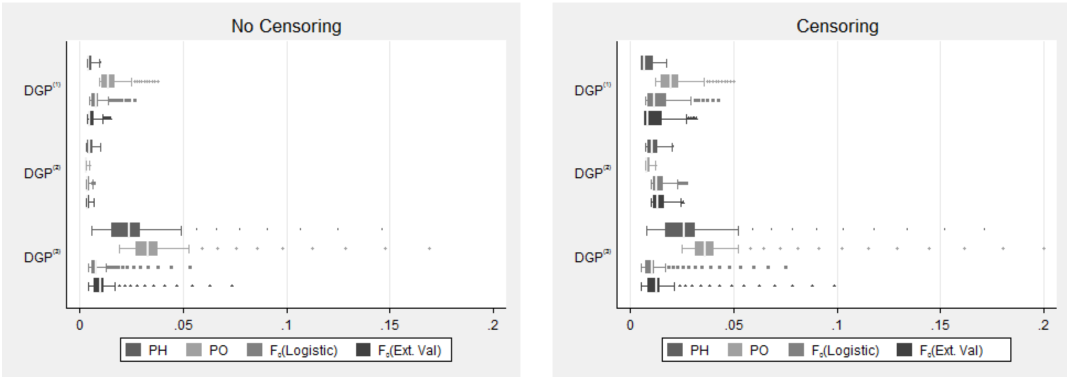
Table 2.5 Maximum Distance (DH) Classical Survival Models and DR

	Cens. level	Sample size	PH	PO	$\mathbf{F_0}$(Logistic)	$\mathbf{F_0}$(Ext.value)
$DGP^{(1)}$	**0%**	**100**	83.53	85.39	97.37	95.15
		500	39.15	43.61	51.10	46.25
		1500	22.47	30.76	34.74	29.89
	30%	**100**	97.15	119.85	120.07	120.40
		500	45.69	49.74	59.49	55.08
		1500	26.89	35.43	39.92	33.20
$DGP^{(2)}$	**0%**	**100**	80.85	81.31	90.93	90.49
		500	37.52	37.51	50.27	50.93
		1500	21.73	21.55	44.95	46.31
	30%	**100**	94.45	104.27	118.42	118.60
		500	44.53	44.66	52.95	52.87
		1500	25.89	25.97	32.95	33.45
$DGP^{(3)}$	**0%**	**100**	117.01	121.56	104.81	111.21
		500	99.93	105.10	66.18	77.24
		1500	93.90	98.63	57.55	69.64
	30%	**100**	122.50	154.16	122.72	130.64
		500	92.24	97.04	69.53	80.13
		1500	82.97	88.41	56.65	68.80

Note: Conditional distribution is evaluated at $x = 0.5$. DR estimates are rearranged using DH algorithm. MD multiplied by 1000 to facilitate comparisons.

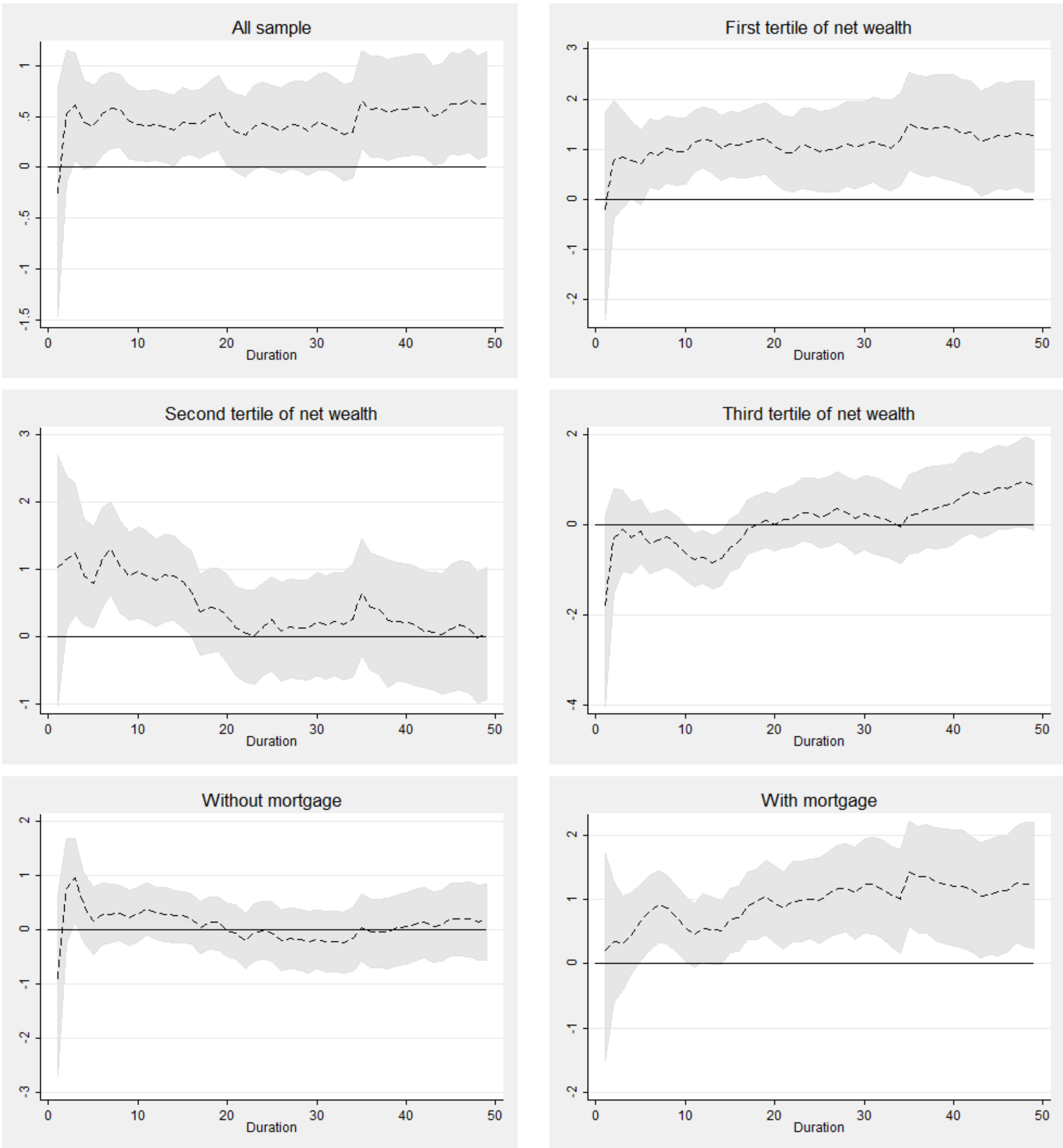
2.8.3 Figures

Figure 2.2 Comparison Classical Survival Models and DR



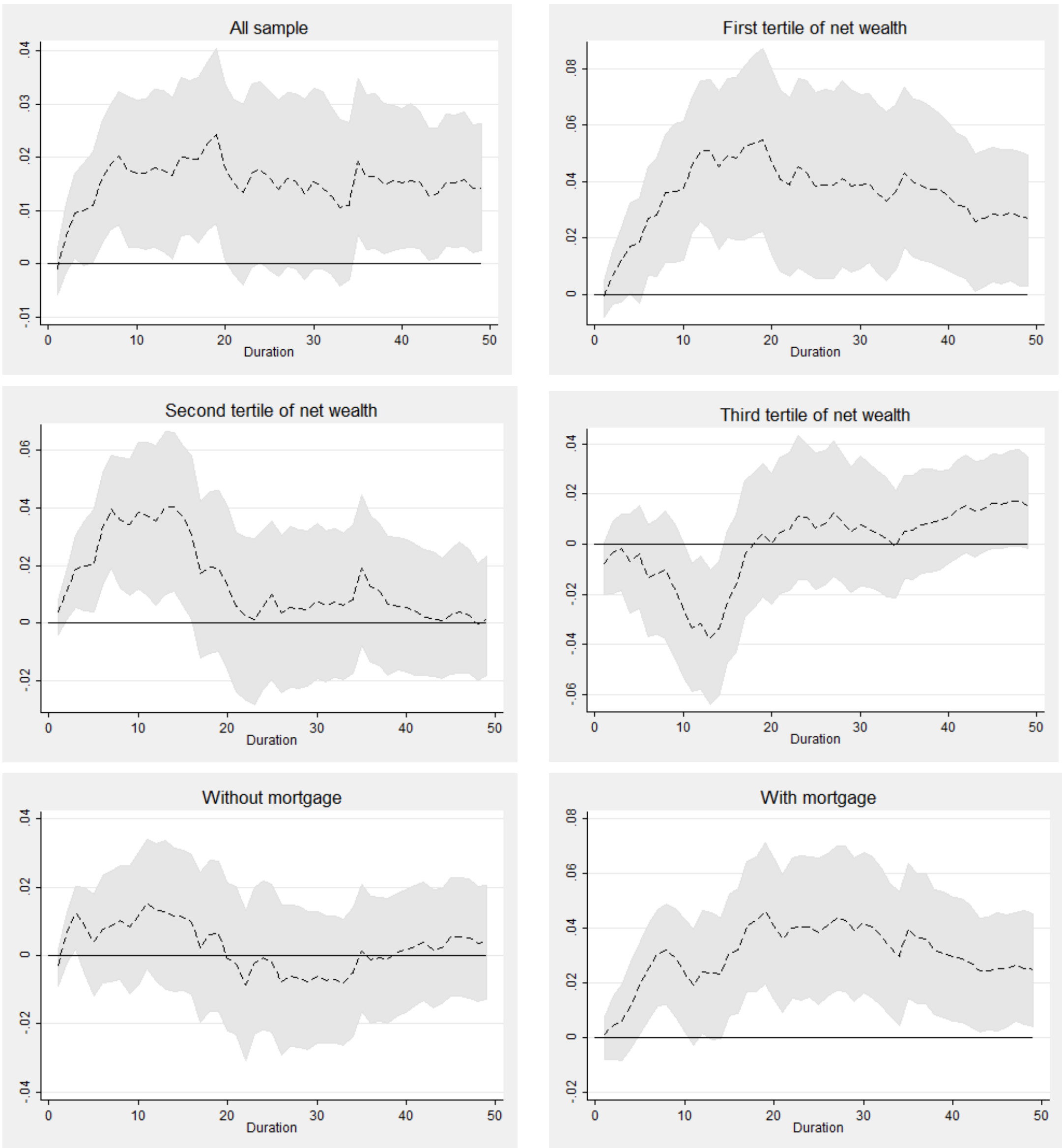
Note: Authors' calculations.

Figure 2.3 DR Coefficients of Unemployment Benefits on Unemployment Duration



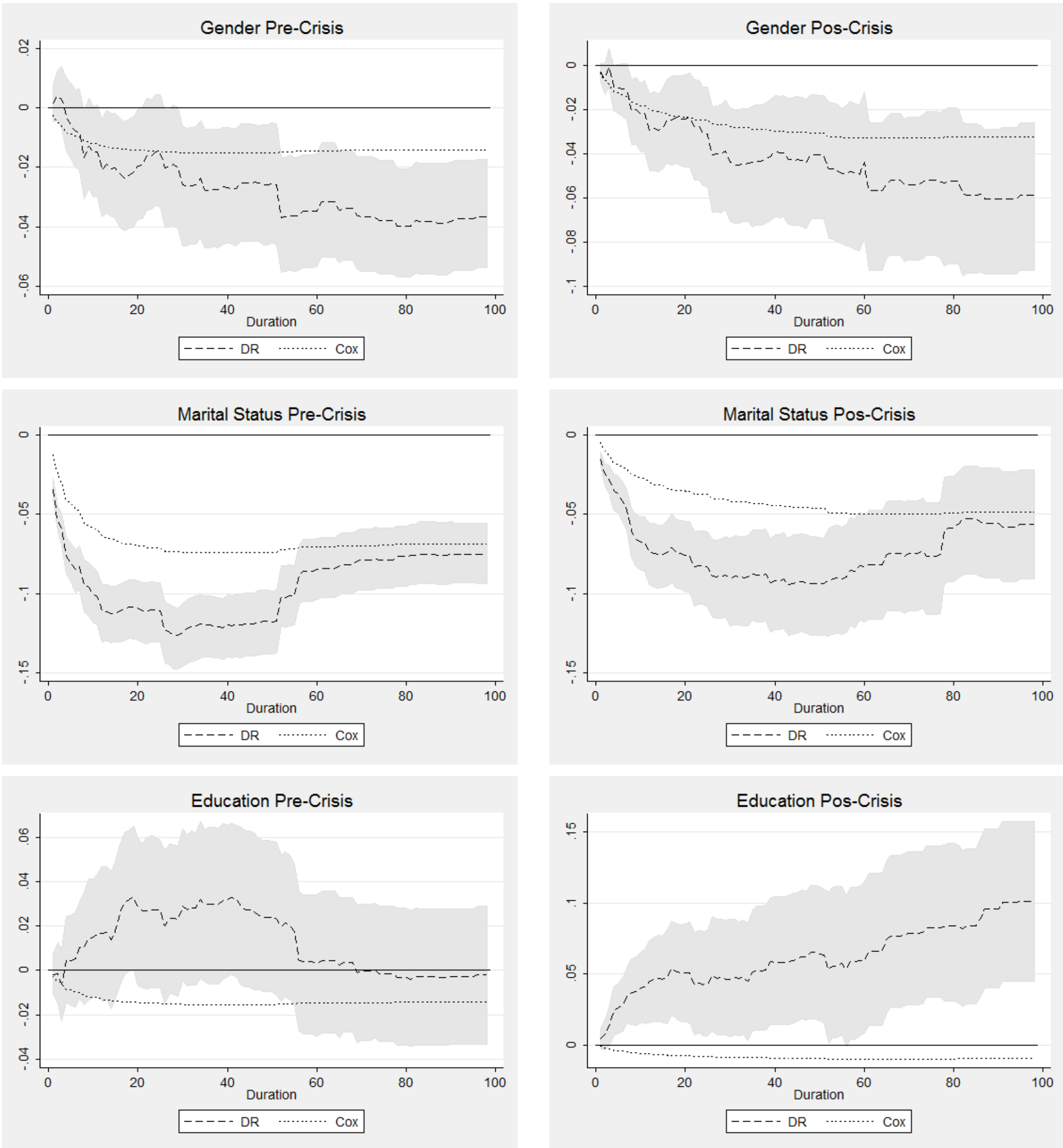
Note: Authors' calculations. 90% confidence intervals computed by bootstrapping.

Figure 2.4 Marginal Effects of Unemployment Benefits on Unemployment Duration Distribution



Note: Authors' calculations. 90% confidence intervals computed by bootstrapping.

Figure 2.5 Marginal Effects on Survival Distribution



Note: Authors' calculations. 90% confidence intervals computed by bootstrapping.

3 Inference on Survival Econometric Models Using Censored Data

3.1 Introduction

This chapter adapts the generalized method of moments (GMM) to estimating parameters identified by moment restrictions involving survival time observed under right random censoring. This offers an alternative to survival analysis based on econometric modelling rather than assuming a particular parametric form for the conditional hazard function, which is the classical statistical inference tool.

The most popular econometric modelling strategy consists on specifying a linear in parameters structural model relating the target variable and covariables, where parameters are identified by means of orthogonality conditions between errors and a vector of instrumental variables. Then, parameters are estimated by the method of moments. This method was proposed by Pearson (1894) and generalized by Neyman and Pearson (1928) by exploiting overidentification restrictions, when more restrictions than parameters are available.

The GMM minimizes in the parameters the sample analog of the quadratic form of moment restrictions and a weighting matrix, which can be chosen in an optimal way to maximize efficiency. Hansen (1982) developed the asymptotic properties of GMM estimators in a general set up. See the monographs of Cochrane (2001); Hall and Horowitz (2005); Singleton (2006) and the reviews by Ogaki (1993) and Hansen (2001) for further discussion. These classical econometric techniques designed to make inferences on structural models are not suitable when some variable is observed under random censoring.

Estimation based on instrumental variables like two and three stages least squares for linear homoskedastic models, nested in the GMM class, has not been yet adapted to the estimation of structural models involving survival time observed under censoring. For instance, returns of education in terms of unemployment duration have been modelled by specifying a conditional hazard function. This approach involves unafordable restrictions on the structure of the marginal effect of education, e.g. that education is exogenous, covariables do not depend on time or the conditional distribution of unemployment duration is of known parametric or semiparametric form. In this context, it seems more natural to consider a structural model relating unemployment duration and education with parameters identified using instrumental variables, mimicing classical models relating wages equations to estimate education returns.

For the sake of motivation, consider a $\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^k - valued$ random vector (T, C, X) defined on $(\Omega, \mathcal{F}, \mathbb{P})$, where T is survival time, C is a censoring variable and X is a set of covariates that may depend on T. Under censoring, we only observe (Y, δ, X), where $Y = \min(T, C)$ and $\delta = 1_{\{T \leq C\}}$, where $1_{\{.\}}$ is the indicator function of whether T is censored. In the context of survival analysis, T might be observed under censoring due to lack of follow-up. In particular, collecting duration data requires following individuals over time and it might occur that individuals either do not change their status during the follow-up period or withdraw.

Classical survival models, usually consider a functional form for the conditional hazard functions, e.g. the proportional hazard model (Cox, 1972, 1975), the propotional odds model (Clayton, 1976) or the accelarated failure model (Kalbfleisch and Prentice, 1980). All these models can be expressed in terms of a transformed survival time linear regression model, where the error distribution is known (see Doksum and Gasko, 1990 for further discussion). That is, consider two subvectors of X, say Z and V, which may share some components. A classical conditional hazard model of T given the $\mathbb{R}^p-valued$ vector Z can be expressed as the transformed regression model

$$\varrho(T) = Z'\beta_0 + U \ a.s., \tag{3.1}$$

where "$'$" means transpose, $\varrho : \mathbb{R}^+ \to \mathbb{R}$ is a monotonically increasing function, β_0 an unknown vector of parameters, and U an error term. Moreover, it is assumed that U follows a known distribution and is independent of Z, which is hard to afford in practice.

The modelling strategy in econometrics consists of exploiting out-of-sample information, coming from economic theory or otherwise, which justifies that there exists a $\mathbb{R}^m-$valued random vector V, called vector of instruments, such that $\mathbb{E}(VU) = 0$. The parameter vector is identified if $m \geq p$ and $rank\left[\mathbb{E}(ZV')\right] = p$.

In the previous example on returns of education where a target variable is unemployment duration, a reasonable transformation is $\varrho(t) = \log t$, which results a natural interpretation of parameters in terms of elasticities. However, the parameter associated to education in Equation (3.1) may not capture the causal effect on unemployment duration due to expected correlation between education and U. Such correlation might be caused by the presence of unobservable factors correlated with unemployment status and education, e.g. networks or innate ability. Despite the usual sources of endogeneity we can expect that education and errors depend on time. We may also consider models where T is a explanatory variable. For instance, we may be interested in estimating the marginal effect of unemployment duration on reemployment wages or some major life event such that the probability of getting divorce.

This chapter justifies inferences for model (3.1) based on instrumental variables, as well as general structural nonlinear models based on the GMM. Under certain restrictions on the relation between the survival time T and the censoring variable C, the joint distribution of (T, X) can be identified from the observed random vector (Y, X, δ), which forms a basis to consistently estimate moment restrictions involving (T, X) and, hence, the parameters of the model using GMM. We provide sufficient conditions for consistency and asymptotic normality of the GMM estimators and a consistent estimator of the optimal weighting matrix.

The rest of the chapter is organized as follows. In next section, we discuss the estimation of parameters identified by means of a set of, posibly non-linear in parameters, moment restrictions, which includes the linear model discussed before. Third section we study the finite sample properties through a Monte Carlo simulation exercise of a linear in parameter structural model. We discuss extensions and suggestions for further research in a final section. Mathematical details and further discussion on technical results are presented in the Appendix.

3.2 GMM in Survival Analysis

Consider a $m \times 1$ vector of functions $\eta_\theta : \mathbb{R}^{1+k} \to \mathbb{R}^m$ indexed by a parameter $\theta \in \Theta \subseteq \mathbb{R}^p$, $m \geq p$, and define

$$\Psi(\theta) := \mathbb{E}\left[\eta_\theta(T, X)\right] = \int_{\mathbb{R}^{1+k}} \eta_\theta(t, x) F(dt, dx), \tag{3.2}$$

where F is the joint distribution function of (T, X), and the $m \times p$ Jacobian matrix $\dot{\Psi}(\theta) := \partial \Psi(\theta) / \partial \theta'$. The parameter $\theta_0 \in \Theta$ is the only solution to the system of equations

$$\Psi(\theta_0) = 0. \tag{3.3}$$

Consider $\{T_i, C_i, X_i\}_{i=1}^n$ $i.i.d$ as (T, C, X), and suppose by the moment that T

is observed. In this case, $\Psi(\theta)$ can be estimated by its empirical analog,

$$\hat{\Psi}(\theta) = \int_{\mathbb{R}^{1+k}} \eta_\theta(t,x)\hat{F}(dt,dx) = \sum_{i=1}^{n} \frac{1}{n}\eta_\theta(T_i,X_i) \tag{3.4}$$

which results of replacing F by the empirical distribution, i.e.,

$$\hat{F}(t,x) = \sum_{i=1}^{n} \frac{1}{n} 1_{\{T_i \le t, X_i \le x\}}. \tag{3.5}$$

A natural estimator of θ_0 in the exact identified case, $p = m$, is the method of moments, i.e. the empirical version of θ_0 in (3.3) that solves $\hat{\Psi}(\theta) = 0$. In the overidentified case, i.e. when $m > d$, there is no a unique $\theta \in \Theta$ such that $\hat{\Psi}(\theta) = 0$, then the GMM estimator makes $\hat{\Psi}(\theta)$ as closed as possible to 0 according to the norm $\|a\|_A = \sqrt{a'Aa}$ for vector a and a given positive definite matrix A, which may depend on data. When $A = I$, where I is the identity matrix, $\|a\|_I$ is the Euclidean distance. Hence, the GMM estimator using a matrix A is

$$\hat{\theta}(A) = \arg\min_{\theta \in \Theta} \left\| \hat{\Psi}(\theta) \right\|_A^2. \tag{3.6}$$

The usual sufficient regularity conditions for consistency are as follows,

Assumption 3.1 *Suppose following conditions hold:*

a. A is p.d.

b. Θ is compact.

c. $\eta_\theta(Y,X)$ is Borel and continuous at each $\theta \in \Theta$ a.s..

d. $\mathbb{E}\left[\sup_{\theta\in\Theta} \|\eta_\theta(Y,X)\|_I\right] < \infty$.

e. Equation (3.3) *holds and $A\Psi(\theta) \neq 0$ for all $\theta \neq \theta_0$.*

Condition 3.1.a guarantees the existence of the estimator. Conditions 3.1.b-d are sufficient for the *a.s.* uniform convergence of $\hat{\Psi}(\theta)$ to $\Psi(\theta)$ (e.g. Jennrich, 1969 1969, Theorem 2 or Newey and McFadden, 1994, Lemma 2.4), and 3.1.e is sufficient for the global identifiability of θ_0 (see c.f. Lemma 2.3 in Newey and McFadden, 1994). For asymptotic normality, further smoothness conditions are needed (Newey and McFadden, 1994, Theorem 3.4).

Assumption 3.2 *Suppose following conditions hold:*
a. θ_0 *is an interior point of* Θ.
b. $\mathbb{E}\left[\left\|\eta_{\theta_0}(Y,X)\right\|_I^2\right] < \infty$.
c. $\eta_\theta(Y,X)$ *is continuously differentiable in a neighborhood of* θ_0 *a.s.*
d. $rank(\dot{\Psi}(\theta_0)) = p$, *where* $\dot{\Psi}(\theta) = \left[\frac{\partial}{\partial\theta_1}\Psi(\theta), ..., \frac{\partial}{\partial\theta_p}\Psi(\theta)\right]$ *is the Jacobian of* $\Psi(\theta)$.
e. $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\partial\eta_\theta(Y,X)/\partial\theta_j\right\|_I\right] < \infty$, $j = 1, ..., p$

These conditons are sufficient to apply the central limit theorem (CLT) for $\hat{\Psi}(\theta_0)$, and guarantees that $\Gamma_{\theta_0}(A) = \dot{\Psi}(\theta_0)A\dot{\Psi}(\theta_0)'$ is *p.d.* for any *p.d.* matrix A. Notice that Equation (3.3) and Assumption 3.2.d imply local identifiability (c.f. Rothenberg, 1971). Under Assumptions 3.1 and 3.2, $\hat{\theta}(A)$ is consistent and asymptotically normal (CAN) with asymptotic variance matrix

$$AsyVar\left(\hat{\theta}(A)\right) = \frac{1}{n}\Gamma_{\theta_0}(A)^{-1}\dot{\Psi}(\theta_0)A\Omega_0A'\dot{\Psi}(\theta_0)'\Gamma_{\theta_0}(A)^{-1\prime}, \tag{3.7}$$

where

$$\Omega_0 = \mathbb{E}\left[\eta_{\theta_0}(T,X)\eta_{\theta_0}(T,X)'\right].$$

The optimal choice of weighting matrix A in order to maximize efficiency is Ω_0^{-1}, which can be estimated given a preliminary θ_0 estimator $\hat{\theta}(A_0)$ based on some initial weighting matrix A_0. That is, Ω_0 is estimated by $\hat{\Omega} = \hat{\Omega}\left(\hat{\theta}(A_0)\right)$, with

$$\hat{\Omega}(\theta) = \int_{\mathbb{R}^{1+k}} \eta_\theta(t,x)\eta_\theta(t,x)'\hat{F}(dt,dx) \qquad (3.8)$$
$$= \sum_{i=1}^{n} \frac{1}{n}\eta_\theta(T_i,X_i))\eta_\theta(T_i,X_i))',$$

Then, under Assumptions 3.1 and 3.2,

$$AsyVar\left(\hat{\theta}\left(\hat{\Omega}^{-1}\right)\right) = \frac{1}{n}\Gamma_{\theta_0}(\Omega_0^{-1})^{-1},$$

which is the minimal asymptotic variance among the estimators $\hat{\theta}\,(A)$. See Hansen (1982) for discussion.

In survival analysis T is not observed, and inferences must be based on the observed sample $\{Y_i, X_i, \delta_i\}_{i=1}^n$, with $Y_i = \min(T_i, C_i)$ and $\delta_i = 1_{\{T_i \leq C_i\}}$. The estimator $\hat{\theta}(A)$ is unfeasible, but Equation (3.4) suggests to estimate $\Psi(\theta)$ in (3.2) by plugging in a consistent estimator of F based on the available data. To this end, some out of sample information is needed to identify F from the joint distribution of the observed random vector (Y, X, δ).

In the absence of covariates, Kaplan and Meier (1958) found out that the survival time distribution can be identified under random censoring, i.e. assuming that survival time and the censored variable are independent. Stute (1993) provided the extension to the multivariate case, including covariates, which needs an extra restriccion of the dependence of C and X in order to identify F, as stated below.

Assumption 3.3 *The following conditions hold:*
a. $\mathbb{P}(T \leq t, C \leq c) = \mathbb{P}(T \leq t)\,\mathbb{P}(C \leq c)$.

b. $\mathbb{P}(T \leq C|T, X) = \mathbb{P}(T \leq C|T)$.

Assumption 3.3.a. guarantees identification of the marginal distribution of survival times. In turn, Assumption 3.3.b. states the relation between the censoring mechanism and the covariates so that, given the actual survival times T, the covariates do not provide any further information on whether censoring occurs. Next proposition states the main result on identification of F in terms of the subdistributions of (Y, X, δ), $H_d(t, x) = \mathbb{P}(Y \leq t, X \leq x, \delta = d)$, which can be estimated from the data. To this end, we introduce the function

$$\Lambda(t, x) = \int_0^t \frac{F(d\bar{t}, x)}{1 - F(\bar{t}-, \infty)}, (t, x) \in \mathbb{R}^{1+k}, \tag{3.9}$$

where for a generic function J, $J(z-) = \lim_{v \uparrow z} J(v)$ and $\infty = (\infty, ..., \infty)'$. From the multiplicative relation between the survival function and the agregate hazard function,

$$1 - F(t, x) = \exp(-\Lambda^c(t, x)) \prod_{\bar{t} \leq t} [1 - \Lambda(\{\bar{t}\}, x)], \tag{3.10}$$

where $\Lambda^c(., x)$ is the continuous part of $\Lambda(., x)$, and for any generic fucntion J, $J(\{v\}) = J(v) - J(v-)$. Define $H(t-) = H_0(t-, \infty) + H_1(t-, \infty) = \mathbb{P}(Y \leq t)$.

Proposition 3.1 *Under Asumption 3.3,*

$$\Lambda(t, x) = \int_0^t \frac{H_1(d\bar{t}, x)}{1 - H(\bar{t}-)}.$$

Then $H_d(t, x)$ can be estimated by its sampling distribution

$$\hat{H}_d(t, x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{Y_i \leq t,\ X_i \leq x,\ \delta_i = d\}},$$

and, Following Garcia-Suaza (2016), the corresponding estimator of Λ is

$$\hat{\Lambda}(t,x) = \int_0^t \frac{\hat{H}_1(d\bar{t},x)}{1-\hat{H}(\bar{t}-)} = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{Y_i \le t, X_i \le x, \delta_i = 1\}}}{1-\hat{H}(Y_i-)} = \sum_{i=1}^{n} \frac{1_{\{Y_i \le t, \delta_i=1\}}}{n-R_i+1} 1_{\{X_i \le x\}},$$

where R_i is the rank of Y_i among $Y_1, ..., Y_n$. According to (3.10),

$$1-\hat{F}(t,x) = \prod_{\bar{t} \le t} \left[1-\hat{\Lambda}(\{\bar{t}\},x)\right],$$

which is the extension of the Kaplan and Meier (1958) estimator. The jump size equals $\hat{\Lambda}(\{Y_i\},x) = [n-R_i+1]^{-1} 1_{\{X_i \le x\}}$, and hence,

$$1-\hat{F}(t,x) = \prod_{Y_i \le t, X_i \le x, \delta_i = 1} \frac{n-R_i}{n-R_i+1} = \prod_{Y_i \le t, X_i \le x} \left(1 - \frac{\delta_i}{n-R_i+1}\right).$$

After ordering the $Y_i's$, we obtain,

$$1-\hat{F}(t,x) = \prod_{Y_{i:n} \le t, X_{[i:n]} \le x} \left(1-\frac{\delta_{[i:n]}}{n-i+1}\right),$$

where $Y_{1:n} \le ... \le Y_{n:n}$ and for generic random vectors $\{\xi_i\}_{i=1}^n$, $\xi_{[i:n]}$ is the i-th $\xi-$concomitant of $Y_{i:n}$, *i.e.* $\xi_{[i:n]} = \xi_j$ if $Y_{i:n} = Y_j$. Therefore, the mass given to $(Y_{i:n}, x)$ is, for each $x \in \mathbb{R}^k$, the size of the jump from $\left(Y_{i-1:n}, X_{[i-1:n]}\right)$ to $\left(Y_{i:n}, X_{[i:n]}\right)$ is,

$$W_i = \hat{F}\left(Y_{i:n}, X_{[i:n]}\right) - \hat{F}_n\left(Y_{i-1:n}, X_{[i-1:n]}\right) = \frac{\delta_{[i:n]}}{n-i+1} \prod_{j=1}^{i-1} \left[1-\frac{\delta_{[j:n]}^{(\ell)}}{n-j+1}\right]. \tag{3.11}$$

In case of ties, it is considered as if uncensored observations precede the censored observations, and other kind of ties are ordered arbitrarily. Noticeably, in the absence of censoring, and $\{T_i, X_i\}_{i=1}^n$ are observed, $W_i = n^{-1}$ and $\hat{F}$ is the sample distribution. Finally, using 3.11, we have that the joint distribution can be express as follows (see Garcia-Suaza, 2016 for details),

$$\hat{F}(t,x) = \sum_{i=1}^{n} W_i 1_{\left\{Y_{i:n} \le t,\ X_{[i:n]} \le x\right\}}.$$

Therefore, a natural estimator of $\Psi(\theta)$ is obtained by plugging-in $\hat{F}$ in Equation (2.3), that is,

$$\hat{\Psi}(\theta) = \int_{\mathbb{R}^{1+k}} \eta_\theta(t,x)\hat{F}(dt,dx) = \sum_{i=1}^{n} W_i \eta_\theta(Y_{i:n}, X_{[i:n]}).$$

Next Proposition applies Stute (1993, 1999) results to obtain the probabilistic limit of $\hat{\Psi}(\theta)$. This estimator may be biased without imposing further assumptions on the support F. Let $\tau_H = \inf\{t : H(t) = 1\}$ be the least upper bound for the support of the marginal distribution of Y, and consider the analogous of T and τ_F and τ_G, with $G(t) = \mathbb{P}(C \le t)$.

Proposition 3.2 *Under Assumption 3.1 and 3.3, uniformly in $\theta \in \Theta$,*

$$\hat{\Psi}(\theta) = \int_{\{Y<\tau_H\}} \eta_\theta(Y,X)d\mathbb{P} + 1_{\{\tau_H \in M\}} \int_{\{Y=\tau_H\}} \eta_\theta(\tau_H, X)d\mathbb{P} + o(1) \ a.s.$$

where M is the set of H atoms.

For most lifetime distributions $\tau_H = \infty$. In such situation, according to Proposition 3.2, $\hat{\Psi}(\theta)$ is a consistent estimator of $\Psi(\theta)$ uniformly in $\theta \in \Theta$. But, in the case when $\tau_H = \tau_G < \tau_F$, the inference is restricted to $\left(0, \tilde{Y}\right]$, $\tilde{Y} \le Y_{n:n} < \tau_H$, i.e. the $\hat{\Psi}(\theta)$ estimator is consistent for the truncated moments. See Efron (1967); Meier (1975); Chen et. al. (1982); Mauro (1985) for further discussion. However, there is not need of this asssumption for the consistency of $\hat{\theta}(A)$, as we state in the following Proposition.

Proposition 3.3 *Under Assumption 3.1 and 3.3,*

$$\hat{\theta}_n(A) = \theta_0 + o(1) \ a.s.$$

Under the additional smoothness conditions in Assumption 3.2, $\hat{\theta}(A)$ is CAN, but with a rather different asymptotic variance than in (3.7). Using Stute (1996), we can provide, for each $\theta \in \Theta$, an asymptotic expansion of $\left(\hat{\Psi}(\theta) - \Psi(\theta)\right)$ in terms of the sample average of

$$\psi_\theta(Y, X; \eta) = \eta_\theta(Y, X)\gamma_0(Y)\delta + \gamma_{1\theta}(Y)(1-\delta) - \gamma_{2\theta}(Y),$$

where

$$\gamma_0(t) = \exp\left\{\int_0^{t-} \frac{H_0(d\bar{t}, \infty)}{1 - H(\bar{t})}\right\},$$

$$\gamma_{1\theta}(t) = \frac{1}{1 - H(t)} \int 1_{\{t<\bar{t}\}} \eta_\theta(\bar{t}, \bar{x})\gamma_0(\bar{t})\, dH_1(\bar{t}, \bar{x}),$$

$$\gamma_{2\theta}(t) = \int\int \frac{1_{\{\bar{s}<t, \bar{s}<\bar{t}\}} \eta_\theta(\bar{t}, \bar{x})\gamma_0(\bar{t})}{[1 - H(\bar{s})]^2} H_0(d\bar{s}, \infty)\, H_1(\bar{t}, \bar{x}).$$

In order to justify such expansion, following conditions are required.

Assumption 3.4 *Consider that following integrability conditions hold:*

a. $\int [\eta_\theta(Y, X)\gamma_0(Y)\,\delta]\, d\mathbb{P} < \infty$.

b. $\int |\eta_\theta(Y, X)|\, K^{1/2}(Y)\, d\mathbb{P} < \infty$ *with*

$$K(t) = \int_0^{t-} \frac{G(d\bar{t})}{[1 - H(\bar{t})]\,[1 - G(\bar{t})]}.$$

Condition 3.4.a guarantees that the second moment of the first component of ψ_θ converges, while 3.4.b controls the bias of the Kaplan-Meier integrals (see Stute, 1994 for further discussion). Next proposition follows directly from Stute (1996).

Proposition 3.4 *Under Assumptions 3.1-3.4*

$$\hat{\Psi}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} \psi_{\theta_0}(Y_i, X_i; \eta) + o_p\left(\frac{1}{\sqrt{n}}\right).$$

Therefore, the asymptotic distribution of $\hat{\theta}(A)$ is obtained applying standard GMM theory, where

$$\Sigma_0 = \mathbb{E}\left[\psi_{\theta_0}(Y_i, X_i; \eta)\psi_{\theta_0}(Y_i, X_i; \eta)'\right]$$

takes the role of $\Omega_0 = \mathbb{E}\left[\eta_{\theta_0}(T, X)\eta_{\theta_0}(T, X)'\right]$ in the asymptotic variance of $\hat{\theta}(A)$.

In order to estimate $\Gamma_{\theta_0}\left(\Sigma_0^{-1}\right)$ we need prior estimators of $\gamma_0(t)$, $\gamma_{1\theta}(t)$ and $\gamma_{2\theta}(t)$. Natural estimators are their sample analogs based on the observed data

$$\hat{\gamma}_0(t) = \exp\left\{\int_0^{t-} \frac{\hat{H}_0\left(d\bar{t}, \infty\right)}{1 - \hat{H}(\bar{t})}\right\} = \exp\left\{\frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{Y_i \leq t\}}\delta_i}{1 - \hat{H}(Y_i)}\right\}$$

$$\begin{aligned}
\hat{\gamma}_{1\theta}(t) &= \frac{1}{1 - \hat{H}(t)} \int_{\mathbb{R}^{1+k}} 1_{\{t < \bar{t}\}} \eta_\theta(\bar{t}, \bar{x}) \hat{\gamma}_0\left(\bar{t}\right) d\hat{H}_1\left(\bar{t}, \bar{x}\right) \\
&= \frac{1}{1 - \hat{H}(t)} \frac{1}{n} \sum_{i=1}^{n} 1_{\{t < Y_i\}} \eta_\theta(Y_i, X_i) \hat{\gamma}_0(Y_i) \delta_i,
\end{aligned}$$

$$
\begin{aligned}
\hat{\gamma}_{2\theta}(t) &= \int_{\mathbb{R}^{1+k}} \int_{\mathbb{R}} \frac{1_{\{\bar{s}<t,\bar{s}<\bar{t}\}} \eta_\theta(\bar{y},\bar{x}) \hat{\gamma}_{0n}(\bar{t})}{\left[1-\hat{H}(\bar{s})\right]^2} \hat{H}_0(d\bar{s},\infty) \hat{H}_1(\bar{t},\bar{x}) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1_{\{Y_j<t,Y_j<Y_i\}} \eta_\theta(Y_i,X_i) \hat{\gamma}_{0n}(Y_i) \delta_i (1-\delta_j)}{[1-H(Y_j)]^2}.
\end{aligned}
$$

Therefore, $\psi_\theta(t,x)$ is estimated by,

$$
\hat{\psi}_\theta(t,x;\eta) = \eta_\theta(t,x)\hat{\gamma}_0(t)\delta + \hat{\gamma}_{1\theta}(t)(1-\delta) - \hat{\gamma}_{2\theta}(t).
$$

Proposition 3.5 *Under Assumptions 3.1-3.4*

$$
\sqrt{n}\left(\hat{\theta}(A)-\theta_0\right) \xrightarrow{d} \mathcal{N}_p\left(0, \Gamma_{\theta_0}(A)^{-1} \dot{\Psi}(\theta_0) A \Sigma_0 A' \dot{\Psi}(\theta_0)' \Gamma_{\theta_0}(A)^{-1\prime}\right).
$$

Following Hansen (1982), the asymptotically efficient weighting matrix A is Σ_0^{-1}.

Corollary 3.1 *Under Assumptions 3.1-3.4*

$$
\sqrt{n}\left(\hat{\theta}_n(\Sigma_0^{-1})-\theta_0\right) \xrightarrow{d} N_p\left(0, \Gamma_{\theta_0}\left(\Sigma_0^{-1}\right)^{-1}\right).
$$

Given a preliminary estimator, obtained with a matrix A_0, Σ_0 is estimated by $\hat{\Sigma}^{(0)} = \hat{\Sigma}\left(\hat{\theta}(A_0)\right)$, with,

$$
\hat{\Sigma}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_\theta(Y_i,X_i;\eta) \hat{\psi}_\theta(Y_i,X_i;\eta)'.
$$

The optimality of $\hat{\theta}(\hat{\Sigma}^{-1})$ is stablished in the following Proposition.

Proposition 3.6 *Under Assumptions 3.1-3.4,*

$$\hat{\Sigma} = \Sigma_0 + o_p(1)$$

and

$$\hat{\theta}(\hat{\Sigma}^{-1}) - \hat{\theta}(\Sigma_0^{-1}) = o_p\left(\frac{1}{\sqrt{n}}\right)$$

Therefore, a consistent estimator of $AsyVar\left(\sqrt{n}\hat{\theta}(\hat{\Sigma}^{-1})\right) = \Gamma_{\theta_0}(\Sigma_0^{-1})^{-1}$ is $\hat{\Gamma}_{\hat{\theta}(\hat{\Sigma}^{-1})}(\hat{\Sigma}^{-1})^{-1}$.

3.3 Monte Carlo simulations

To study the finite sample properties of the GMM estimator described above, we perform a simple Monte Carlo simulation in the context of linear models in the just identified case, i.e, one endogenous variable and one instrument. To do so, three different designs are considered: $DGP^{(1)}$ a homoskedastic model, $DGP^{(2)}$ a heteroskedastic model and $DGP^{(3)}$ homoskedastic model with informative censoring. The latter models allows to investigate the performance of the GMM estimator when Assumption 3.3 fails. Therefore, we generate a linear process for T depending on Z and $\tilde{Z}$, where the latter will be ommited in the estimation stage. Therefore, there is an instrument V available such that $\mathbb{E}\left[\tilde{Z}V\right] \neq 0$.

To be specific, we set $T = Z'\theta_0 + U$ where the vector of error depends on $\tilde{Z}$. For $DGP^{(1)}$ and $DGP^{(3)}$, the homoskedastic cases, $U = \tilde{Z} + \varepsilon$ with ε is an *i.i.d* standard normal random variable. While for the heteroskedastic model $DGP^{(2)}$, $U = \tilde{Z} + \varepsilon\,|Z|$. Z, $\tilde{Z}$ and V are generated as uniform random variables such that $Corr\left(Z, \tilde{Z}\right) = 0.5$, $Corr\left(Z, V\right) = 0.5$, and $\tilde{Z}$ is independent of V. In this manner, $\mathbb{E}\left[ZU\right] \neq 0$ but $\mathbb{E}\left[VU\right] = 0$. Finally the target parameter is $\theta_0 = 1$.

Following the strategy in Stute (1993), censoring times follow uniform distribution. Particularly, C is uniform$\left(0, a^{(i)}\right)$ for $DGP^{(1)}$ and $DGP^{(2)}$, and $C = 0.5X+$Uniform$\left(0, a^{(i)}\right)$ for $DGP^{(3)}$. The values $a^{(i)}$ are defined to achieve censoring levels of 0%, 10% and 30%. That is, $a^{(1)} = a^{(2)}$ taking values of 20 and 6.5 for censoring levels of 10% and 30% respectively. In turn, the corresponding values of $a^{(3)}$ are 17 and 6. Finally, we consider 1000 Monte Carlo repetitions and sample size of 100, 300 and 1000.

Since T is not observed, we compute $Y = \min(T, C)$ and $\delta = 1_{\{T \leq C\}}$. In this case, θ_0 is the solution of

$$\Psi(\theta_0) = \mathbb{E}\left[V\left(T - Z'\theta_0\right)\right] = 0.$$

That is, $\eta_\theta = V\left(T - Z'\theta_0\right)$, and hence, $\hat{\theta}$ solves

$$\hat{\Psi}(\theta) = \sum_{i=1}^{n} W_i V'_{i:n}\left(Y_{i:n} - Z'_{[i:n]}\theta\right) = 0.$$

Note that in absence of censoring, $\hat{\theta}$ is the classical intrumental variable estimator. Moreover, because θ_0 is exactly identified, the choice of A plays no role.

Hence, we make comparisons between the classical instrumental variables estimator and the proposed GMM estimator. That is, the estimation when censoring level is 0%, with the case when censoring is present. Such comparison is done in terms of mean bias and root mean squared error (RMSE).

Results reported in Table 3.1 suggest that the GMM procedure in presence of censoring produces estimates as good as in the IV procedure in terms of the average bias and RMSE. As expected, both the mean bias and the RMSE increase as the censoring becomes more severe, but they also reduce with sample size in all designs, including $DGP^{(3)}$. Overall, our results show satisfactory performance of the GMM in a linear model.

Table 3.1 Simulation Results GMM Estimator under Censoring

		Mean bias			RMSE		
	Censoring Level	**100**	**300**	**1000**	**100**	**300**	**1000**
$DGP^{(1)}$	**0%**	-0.0339	-0.0038	-0.0146	0.7891	0.4419	0.2438
	10%	-0.0259	-0.0124	-0.0128	0.8464	0.4655	0.2585
	30%	-0.0280	-0.0075	-0.0116	0.9810	0.5799	0.3129
$DGP^{(2)}$	**0%**	-0.0186	-0.0042	-0.0004	0.4994	0.2721	0.1487
	10%	-0.0327	-0.0116	-0.0014	0.5487	0.2881	0.1580
	30%	-0.0853	-0.0312	-0.0077	0.6845	0.3597	0.2039
$DGP^{(3)}$	**0%**	-0.0294	-0.0077	-0.0076	0.7907	0.4395	0.2419
	10%	-0.0152	0.0000	0.0003	0.8678	0.4706	0.2580
	30%	0.0240	-0.0342	0.0326	1.0627	0.5787	0.3166

3.4 Final Remarks and Further Research

We propose a general GMM estimator which is suitable for randomly censored data. In order to tackle the censoring problem, it is plugging-in a suitable estimator of the joint distribution of the survival times and covariates which makes the produce as simple to compute as in the case of no censoring. In fact, in absence of censoring, the propose estimator is equivalent to the classical GMM procedure.

This GMM estimator facilitates the estimation of causal relationships in the context of survival analysis using similar usual econometric modelling strategies. Therefore, many relevant analysis related to duration outcomes can be conducted. For instance, the estimation of the causal effect of education on unemployment, but also other kind of relations where the duration outcome is a covariate, e.g. the simultaneity of reemployment wages and unemployment duration or the occurence of major life events like divorce or illness during unemployment spell.

The proposed procedure provide the basis for constructing test of practical interest such that a Durbin–Wu–Hausman endogeneity test, a J-test for overidentifying restrictions, and tests for weak instruments. As well as the implementation of analogous procedures to the LIML or the Fulller-k estimators that are more robust against weak instruments. This can be done by replacing the empirical distribution function with the multivariate Kaplan-Meier distribution function.

Lastly, an interesting extension of the GMM procedure consists on regarding time-varying covariates. One of the possible sources of endogeneity is that covariates and unobservables might depend on the underlying duration outcome. For instance, education and skills can be affected by the unemployment duration. Therefore, estimating the causal effect requires to consider all history of covariates.

3.5 Appendix

Proof of Proposition 3.1

This proposition follows the same the arguments of Garcia-Suaza (2016). That is, under Assumption 3.3.b. we have:

$$\begin{aligned} H_1(t,x) &= \mathbb{P}(Y \leq t, X \leq x, \delta = 1) = \mathbb{P}(T \leq t, X \leq x, T \leq C) \\ &= \int_0^t [1 - G(\bar{t}-)] F(d\bar{t}, x), \end{aligned}$$

and by Assumption 3.3.a.

$$1 - H(t) = \mathbb{P}(Y > t) = [1 - F(\bar{t}-, \infty)][1 - G(t)].$$

Thus, using equation (3.9)

$$\Lambda(t,x) = \int_0^t \frac{F(d\bar{t}, x)}{1 - F(\bar{t}-, \infty)} = \int_0^t \frac{H_1(d\bar{t}, x)}{1 - H(\bar{t}-)}.$$

Proof of Proposition 3.2

Because of Assumption 3.1.c-d, η_θ is integrable. Then, by Theorem in Stute (1993), with probability one,

$$\lim_{n\to\infty} \hat{\Psi}(\theta) = \int_{\{Y<\tau_H\}} \eta_\theta(Y,X)d\mathbb{P} + 1_{\{\tau_H \in M\}} \int_{\{Y=\tau_H\}} \eta_\theta(\tau_H, X)d\mathbb{P}.$$

Moreover, if $\tau_H = \infty$,

$$\hat{\Psi}(\theta) = \int \eta_\theta(Y,X)d\mathbb{P} + o(1) \ a.s.$$

Proof of Proposition 3.3

By compactness of Θ, continuity and boundedness of η_θ (Assumption 3.1.b-d) and Theorem 2 in Jennrich (1969), $\hat{\Psi}(\theta)$ converges uniformly in θ, that is,

$$\sup_{\theta\in\Theta}\left\|\hat{\Psi}(\theta)-\Psi(\theta)\right\|=0$$

We know that

$$\hat{\theta}(A)=\arg\min_{\theta\in\Theta}\left\|\hat{\Psi}(\theta)\right\|_A^2,$$

for a given A. Since A is *p.d.*, $\left\|\hat{\Psi}(\tilde{\theta})\right\|_{A,n}^2\geq 0$ for every $\tilde{\theta}\in\Theta$ and n, and

$$0\leq\left\|\hat{\Psi}(\hat{\theta})\right\|_{A,n}^2\leq\left\|\hat{\Psi}(\theta)\right\|_{A,n}^2,$$

for every n, $\left\|\hat{\Psi}(\hat{\theta})\right\|_A^2\rightarrow\left\|\hat{\Psi}(\theta)\right\|_A^2$. Consequently, by Assumption 3.1.e, $\hat{\Psi}(\hat{\theta})\neq\hat{\Psi}(\tilde{\theta})$ for $\hat{\theta}\neq\tilde{\theta}$, then $\hat{\theta}_n(A)\rightarrow\theta_0$.

Proof of Proposition 3.4

This result follows from integrability of η_θ, Lemma 5.1 and Theorem 1.1 in Stute (1993).

Proof of Proposition 3.5

Denote

$$\dot{\hat{\Psi}}(\theta)=\frac{\partial\hat{\Psi}(\theta)}{\partial\theta'}$$

which converges in probability from Assumption 3.2.d, i.e. $\dot{\hat{\Psi}}(\theta)\rightarrow\dot{\Psi}(\theta)$. By 3.6, $\hat{\theta}(A)$ solves

$$\dot{\hat{\Psi}}(\hat{\theta})'A\hat{\Psi}(\hat{\theta})=0. \tag{3.12}$$

The first order Taylor expansion around the $j-th$ element of θ_0, θ_{j0}, is given by

$$\hat{\Psi}(\hat{\theta}) = \hat{\Psi}(\theta_0) + \sum_{i=1}^{p} \dot{\hat{\Psi}}_j(\tilde{\theta}_{(i)}) \left(\hat{\theta}_j - \theta_{j0}\right),$$

where $\hat{\theta}_j$ is the $j-th$ element of $\hat{\theta}$, $\dot{\hat{\Psi}}_j$the $j-th$ row of $\dot{\hat{\Psi}}$, and $\tilde{\theta}_{(i)}$ denotes the $p\times 1$ vector such that $\left\|\tilde{\theta}_{(i)} - \theta_0\right\| \leq \left\|\hat{\theta} - \theta_0\right\|$. Since $\hat{\theta} \to \theta_0$, $\tilde{\theta}_{(i)} \to \theta_0$. By replacing in 3.12, we have

$$\dot{\hat{\Psi}}(\hat{\theta})' A \hat{\Psi}(\theta_0) + \dot{\hat{\Psi}}(\hat{\theta})' A \dot{\hat{\Psi}}(\tilde{\theta}) \left(\hat{\theta} - \theta_0\right) = 0.$$

Therefore,

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = -\left[\dot{\hat{\Psi}}(\hat{\theta})' A \dot{\hat{\Psi}}(\tilde{\theta})\right]^{-1} \dot{\hat{\Psi}}(\hat{\theta})' A \sqrt{n} \hat{\Psi}(\theta_0).$$

By Assumption 3.2.c-d, consistency of $\hat{\theta}$, and Theorem in Stute (1993),

$$\dot{\hat{\Psi}}(\hat{\theta}) \to \dot{\Psi}(\theta), \quad \dot{\hat{\Psi}}(\tilde{\theta}) \to \dot{\Psi}(\theta),$$

and hence,

$$\dot{\hat{\Psi}}(\hat{\theta})' A \dot{\hat{\Psi}}(\tilde{\theta}) \to \dot{\Psi}(\theta)' A \dot{\Psi}(\theta) := \Gamma_\theta\left(A\right),$$

and Thereom 1.2 in Stute (1996),

$$\sqrt{n}\left(\hat{\Psi}(\theta) - \Psi(\theta)\right) \xrightarrow{d} N_p\left(0, \Sigma\right),$$

with $\Sigma = \mathbb{E}\left[\psi_\theta(Y, X; \eta)\psi_\theta(Y, X; \eta)'\right]$. Hence,

$$\sqrt{n}\left(\hat{\theta}(A) - \theta_0\right) \xrightarrow{d} N_p\left(0, \Gamma_{\theta_0}\left(A\right)^{-1} \dot{\Psi}(\theta_0) A \Sigma_0 A' \dot{\Psi}(\theta_0)' \Gamma_{\theta_0}\left(A\right)^{-1\prime}\right).$$

Proof of Corollary 3.1

Hansen (1982) shows that the weight matrix that minimizes the variance of the GMM estimator is $A = \Sigma_0^{-1}$. In this case, the asymptotic variance reduces to

$$\begin{aligned} AsyVar\left(\hat{\theta}(A)\right) &= \frac{1}{n}\Gamma_{\theta_0}(A)^{-1}\dot{\Psi}(\theta_0)A\Sigma_0A'\dot{\Psi}(\theta_0)'\Gamma_{\theta_0}(A)^{-1\prime} \\ &= \frac{1}{n}\Gamma_{\theta_0}\left(\Sigma_0^{-1}\right)^{-1} \end{aligned}$$

Proof of Proposition 3.6

This result follows from Theorem in Stute (1993) and consistency of $\hat{\theta}(A)$ (Proposition 3.3).

References

Addison, J.T., Portugal, P., 2003. Unemployment Duration Competing and Defective Risks. Journal of Human Resources 38, 157–191.

Akritas, M.G., 1994. Nearest Neighbor Estimation of a Bivariate Distribution Under Random Censoring. The Annals of Statistics 22, 1299–1327.

Amemiya, T., 1985. Advanced econometrics. Harvard university press.

Andersen, P.K., Gill, R.D., 1982. Cox's Regression Model for Counting Processes: a Large Sample Study. The Annals of Statistics 10, 1100–1120.

Angrist, J., Chernozhukov, V., Fernández-val, I., 2006. Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure. Econometrica 74, 539–563.

Azarang, L., Uña-Álvarez, J.D., Stute, W., 2015. The Jackknife Estimate of Covariance of two Kaplan-Meier Integrals with Covariables. Statistics: A Journal of Theoretical and Applied Statistics 49, 1005–1025.

Azmat, G., Güell, M., Manning, A., 2006. Gender Gaps in Unemployment Rates in OECD Countries. Journal of Labor Economics 24, 1–37.

Bachmann, R., Sinning, M., 2016. Decomposing the Ins and Outs of Cyclical Unemployment. Oxford Bulletin of Economics and Statistics. Forthcoming

Bailey, K.R., 1983. The Asymptotic Joint Distribution of Regression and Survival Parameter Estimates in the Cox Regression Model. The Annals of Statistics 11, 39–48.

Bailey, K.R., 1984. Asymptotic Equivalence Between the Cox Estimator and the General ML Estimators of Regression and Survival Parameters in the Cox Model. The Annals of Statistics 12, 730–736.

Bang, H., Tsiatis, A. a, 2002. Median Regression with Censored Cost Data. Biometrics 58, 643–649.

Bauer, T.K., Sinning, M., 2008. An Extension of the Blinder-Oaxaca Decomposition to Non-Linear Models. Advances in Statistical Analysis 92, 197–206.

Baussola, M., Mussida, C., Jenkins, J., Penfold, M., 2015. Determinants of the Unemployment Gender Gap: A Comparative Investigation. nternational Labour Review 154, 537–562.

Beran, R., 1981. Nonparametric Regression with Randomly Censored Survival Data. Mimeo

Bennett, S., 1983. Analysis of Survival Data by the Proportional Odds Model. Statistics in Medicine 2, 273–277.

Biewen, M., Wilke, R.A., 2004. Unemployment Duration in West-Germany: Do the IAB Employment Subsample and the German Socio-Economic Panel Yield the Same Results?. Unpublished manuscript, University of Frankfurt, University of Frankfurt.

Bitler, M.P., Gelbach, J.B., Hoynes, H.W., 2006. What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments. American Economic Review 96, 988–1012.

Blau, F.D., Kahn, L.M., 1992. The Gender Earnings Gap: Learning from International Comparisons. American Economic Review 82, 533–538.

Blinder, A.S., 1973. Wage Discrimination: Reduced Form and Structural Estimates. Journal of Human Resources 8, 436–455.

Borooah, V.K., 2002. A Duration-Sensitive Measure of the Unemployment Rate: Theory and Application. Labour 16, 453–468.

Bouaziz, O., Lopez, O., 2010. Conditional Density Estimation in a Censored Single-Index Regression Model. Bernoulli 16, 514–542.

Breslow, N., 1974. Covariance Analysis of Censored Survival Data. Biometrics 30, 89–99.

Buckley, J., James, I., 1979. Linear Regression with Censored Data. Biometrika 66, 429–436.

Burr, D., 1994. A Comparison of Certain Bootstrap Confidence Intervals in the Cox Model. Journal of the American Statistical Association 89, 1290–1302.

Cain, G., 1986. The Economic Analysis of Labor Market Discrimination: A Survey, in: Ashenfelter, O and Layard, R. (Ed.), Handbook of Labor Economics. Elsevier Science Publishers.

Cartwright, N., Reiss, J., 2004. Uncertainty in Econometrics: Evaluating Policy Counterfactuals, in: Economic Policy Under Uncertainty: The Role of Truth and Accountability in Policy Advice. 1–29.

Chen, Y.Y., Hollander, M., Langberg, N., 2011. Small-Sample Results for the Estimator. Journal of the American Statistical Association 77, 141– 144.

Chen, L., Lin, D.Y., Zeng, D., 2010. Attributable Fraction Functions for Censored Event Times. Biometrika 97, 713–726.

Cheng, G., Huang, J.Z., 2010. Bootstrap Consistency for General Semiparametric M-estimation. The Annals of Statistics 38, 2884–2915.

Chernozhukov, V., Fernandez-Val, I., Melly, B., 2013. Inference on Counterfactual Distributions. Econometrica 81, 2205–2268.

Chernozhukov, V., Hansen, C., 2005. An IV Model of Quantile Treatment Effects. Econometrica 73, 245-261.

Chetty, R., 2008. Moral Hazard vs. Liquidity and Optimal Unemployment Insurance. Journal of Political Economy 116, 173–234.

Clayton, A.D.G., 1976. An Odds Ratio Comparison for Ordered Categorical Data with Censored Observations. Biometrika 63, 405–408.

Cochrane, 2001. Asset Pricing. Princeton University Press

Cotton, J., 1988. On the Decomposition of Wage Differentials. The Review of Economics and Statistics 70, 236–243.

Cox, D.R., 1972. Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological) 34, 187–220.

Cox, D., 1975. Partial Likelihood. Biometrika 62, 269–276.

Dabrowska, D.M., 1987. Nonparametric Regression with Censored Survival Time Data. Scandinavian Journal of Statistics 14, 181–197.

Dabrowska, D.M., 1989. Uniform Consistency of the Kernel Conditional Kaplan-Meier Estimate. The Annals of Statistics 17, 1157–1167.

Dawid, A.P., 2000. Causal inference Without Counterfactuals. Journal of the American Statistical Association 95, 407–424.

DiNardo, J., Fortin, N.M., Lemieux, T., 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. Econometrica 64, 1001–1044.

Doksum, K., Gasko, M., 1990. On a Correspondence between Models in Binary Regression and Survival Analysis. International Statistical Review 58, 243–252.

Donald, S.G., Hsu, Y.C., 2014. Estimation and inference for Distribution Functions and Quantile Functions in Treatment Effect Models. Journal of Econometrics 178, 383–397.

Du, F.L., Dong, X.-Y., 2009. Why Women Have Longer Unemplyoment Durations than Men in Post-restructuring Urban China. Cambridge Journal of Economics 33, 233–252.

Efron, B., 1967. The Two Sample Problem with Censored Data, in: Proceedings of the Fifth Berkeley Symposium on Math, Statistics and Probability. 831–853.

Efron, B., 1981. Censored Data and the Bootstrap. Journal of the American Statistical Association 76, 312–319.

Efron, B., 1992. Six Questions Raised by the Bootstrap. Exploring the limits of bootstrap. 99–126.

Efron, B., Tibshirani, R., 1986. Bootstrap Methods for Standard Errors, Confidence Intervals, and other Measures of Statistical Accuracy. Statistical Science 1, 54–75.

Essama-Nssah, B., Lambert, P.J., 2011. Influence Functions for Distributional Statistics, Society for the study of Econoic Inequality, ECINEQ Working Paper Series.

Eusamio, E., 2004. El Diferencial de las Tasas de Paro de Hombres y Mujeres en España (1994-1998), CEMFI.

Firpo, S., Fortin, N.M., Lemieux, T., 2007. Decomposing Wage Distributions Using Recentered Influence Function Regressions, University of British Columbia, University of British Columbia.

Firpo, S., Fortin, N., Lemieux, T., 2009. Unconditional Quantile Regressions. Econometrica 77, 953–973.

Firpo, S., Pinto, C., 2011. Identification and Estimation of Distributional Impacts of Interventions using Changes in Inequality Measures. Mimeo.

Foley, M.C., 1997. Determinants of Unemployment Duration in Russia, Center Discussion Paper.

Foresi, S., Peracchi, F., 1995. The Conditional Distribution of Excess Returns: An Empirical Analysis. Journal of the American Statistical 90, 451–466.

Fortin, N., Lemieux, T., Firpo, S., 2011. Decomposition Methods in Economics, in: Handbook of Labor Economics 4. 1–102.

Frandsen, B.R., Frölich, M., Melly, B., 2012. Quantile treatment effects in the regression discontinuity design. Journal of Econometrics 168, 382–395.

Freeman R.B., 1980. Unionism and the Dispersion of Wages. Industrial and Labor Relations Review 34, 3–23.

Freitag, G., Munk, A., 2005. On Hadamard Differentiability in k-sample Semiparametric Models with Applications to the Assessment of Structural Relationships. Journal of Multivariate Analysis 94, 123–158.

Frölich, M., Melly, B., 2013. Unconditional Quantile Treatment Effects Under Endogeneity. Journal of Business & Economic Statistics 31, 346–357.

Fujita, S., 2011. Effects of Extended Unemployment Insurance Benefits : Evidence from the Monthly CPS. Federal Reserve Bank of Philadelphia. Working Paper.

García-Suaza, A., 2016. Oaxaca-Blinder type Decomposition Methods for Duration Outcomes, Documentos de Trabajo Universidad del Rosario.

Gardeazabal, J., Ugidos, A., 2004. More on Identification in Detailed Wage Decompositions. Review of Economics and Statistics 86, 1034–1036.

Gill, R.D., 1980. Censoring and Stochastic Integrals. Statistica Neerlandica 2, 124.

Gill, R.D., 1984. Understanding Cox's Regression Model: a Martingale Approach. Journal of the American Statistical Association 79, 441–447.

Gill, R.D., Johansen, S., 1990. A survey of Product Integration with a View Towards Application in Survival Analysis. The Annals of Statistics 18, 1501–1555.

Gonzalez-Manteiga, W., Cadarso-Suarez, C., 1994. Asymptotic Properties of a Generalized Kaplan-Meier Estimator with some Applications. Journal of Nonparametric Statistics 4, 65–78.

Gonzalo, M.T., Saarela, J., 2000. Gender Differences in Exit Rates fom Unemployment: Evidence from a Local Finnish Labor Market. Finnish Economic Papers 13, 129–139.

Gorfine, M., Goldberg, Y., Ritov, Y., 2014. A Quantile Regression Model for Failure-time Data, Working Paper.

Guo, S., Zeng, D., 2014. An overview of semiparametric models in survival analysis. Journal of Statistical Planning and Inference.

Hall, P., 1988. Theoretical Comparison of Bootstrap Confidence Intervals. Ann. Stat. 16, 927–953.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer, New York.

Hall, P., Horowitz, J.L., 2005. Nonparametric methods for inference in the presence of instrumental variables. The Annals of Statistics 33, 2904–2929.

Ham, J.C., Svejnar, J., Terrell, K., 1999. Women's Unemployment During Transition. Economics of Transition 7, 47–78.

Hansen, L.P., 1982. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 50, 1029–1054.

Hansen, L. P., 2001. Generalized Method of Moments Estimation: a Time Series Perspective. International Encyclopedia of Social and Behavioral Sciences.

He, X., 1997. Quantile Curves without Crossing. The American Statistician 51, 186–192.

Heuchenne, C., Keilegom, I., 2007. Polynomial Regression with Censored Data based on Preliminary Nonparametric Estimation. Annals of the Institute of Statistical Mathematics 59, 273–297.

Hirose, Y., 2011. Asymptotic Linear Expansion of Profile Likelihood in the Cox Model. Mathematical Methods of Statistics 20, 224–231.

Höfler, M., 2005. Causal Inference Based on Counterfactuals. BMC Medical Research Methodology 5, 1-12.

Honore, B., Khan, S., Powell, J., 2002. Quantile Regression Under Random Censoring. Journal of Econometrics 109, 67–105.

Horvitz, D.G., Thompson, D.J., 1952. A Generalization of Sampling Without Replacement from a Finite Universe. Journal of the American Statistical Association 47, 663–685.

Hothorn, T., Kneib, T., Bühlmann, P., 2014. Conditional Transformation Models. Journal of the Royal Statistical Society. Series B: Statistical Methodology 76, 3–27.

Hunter, D.R., Lange, K., 2002. Computing Estimates in the Proportional Odds Model. Annals of the Institute of Statistical Mathematics 54, 155-168.

Jann, B., 2005. Standard Errors for the Blinder-Oaxaca Decomposition. 3rd Ger. Stata Users' Gr. Meet.

Jann, B., 2008. The Blinder-Oaxaca Decomposition for Linear Regression Models. Stata Journal 8, 453–479.

Jennrich, R.I., 1969. Asymptotic Properties of Non-linear Least Squares Estimators. The Annals of Mathematical Statistics 40, 633–643.

Johnson, J.L., 1983. Sex Differentials in Unemployment Rates: A Case for No Concern. Journal of Political Economics 91, 293–303.

Juhn, C., Murphy, K.M., Pierce, B., 1991. Accounting for the Slowdown in Black-White Wage Convergence, in: Kosters, M.H. (Ed.), Workers and Their Wages. American Enterprise Institute, Washington DC. 107–143.

Juhn, C., Murphy, K.M., Pierce, B., 1993. Wage Inequality and the Rise in Returns to Skill. Journal of Political Economics 101, 410–442.

Kalbfleisch, J., Prentice, R., 1973. Marginal Likelihoods based on Cox's Regression and Life Model. Biometrika 60, 267–278.

Kalbfleisch, J., Prentice, R., 1980. The Statistical Analysis of Failure Time Data. John Wiley & Sons.

Kaplan, E.L., Meier, P., 1958. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53, 457–481.

Koenker, R., Basset, G., 1978. Regression Quantiles. Econometrica 46, 33–50.

Koenker, R., Leorato, S., Peracchi, F., 2013. Distributional vs. Quantile Regression. Working Paper, CEIS Working Paper.

Koul, H., Susarla, V., Ryzin, J. Van, 1981. Regression Analysis with Randomly Right-Censored Data. The Annals of Statistics 9, 1276–1288.

Kuhn, P., Skuterud, M., 2004. Internet Job Search and Unemployment Durations. American Economic Review 94, 218–232.

Leconte, E., Poiraud-Casanova, S., Thomas-Agnan, C., 2002. Smooth Conditional Distribution Function and Quantiles under Random Censorship. Lifetime Data Analysis 8, 229–246.

Lewis, D., 1973. Counterfactuals. Oxford: B. Blackwell.

Lipsitz, S.R., Fitzmaurice, G.M., Molenberghs, G., Zhao, L.P., 1997. Quantile Regression Methods for Longitudinal Data with Drop-outs: Application to CD4 Cell Counts of Patients Infected with the Human Immunodeficiency Virus.Journal of the Royal Statistical Society: Series C (Applied Statistics) 46, 463–476.

Machado, J.F., Mata, J., 2005. Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression. Journal Applied Econometrics 20, 445–465.

Mauro, D., 1985. A Combinatoric Approach to the Kaplan-Meier Estimator. The Annals of Statistics 13, 142–149.

McLain, A.C., Ghosh, S.K., 2011. Nonparametric Estimation of the Conditional Mean Residual Life Function based on Censored Data. Lifetime Data Anal., NC State Department of Statistics. Technical Report 14, 514–532.

Meier, P., 1975. Estimation of a Distribution Function from Incomplete Observations, in: Gani, J. (Ed.), Perspectives in Probability and Statistics. 67–87.

Melly, B., 2005. Decomposition of Differences in Distribution using Quantile Regression. Labour Economics 12, 577–590.

Miller, R., 1976. Least Squares with Censored Regression Data. Biometrika 63, 449–464.

Miller, R., Halpern, J., 1982. Regression with Censored Data. Biometrika 69, 521–531.

Murphy, S.A.S., Rossini, A.J., van der Vaart, A., 1997. Maximum Likelihood Estimation in the Proportional Odds Model. Journal of the American Statistical Association 92, 968–976.

Naes, T., 1982. The Asymptotic Distribution of the Estimator for the Regression Parameter in Cox's Regression Model. Scandinavian Journal of Statistics 9, 107–115.

Newey, W., D. McFadden, 1994. Handbook of Econometrics, vol. 4, chap. Large Sample Estimation and Hypothesis Testing, 2113–2148. Amsterdam: Elsevier.

Neyman, J., Pearson, E.S., 1928. On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. Biometrika 1, 175–240.

Niemi, B., 1974. The Female-Male Differential in Unemployment. Industrial and Labor Relations Review 27, 331–350.

OECD, 2013. http://stats.oecd.org/.

O'Neill, J., O'Neill, D., 2006. What do Wage Differentials Tell us About Labor Market Discrimination?, in: The Economics of Immigration and Social Policy. 293–357. National Bureau of Economic Research.

Oaxaca, R.L., 1973. Male-Female Wage Differentials in Urban Labor Markets. International Economic Review. 14, 693–709.

Oaxaca, R.L., Ransom, M.R., 1999. Identification in Detailed Wage Decompositions. Review of Economics and Statistics 81, 154–157.

Ogaki, M. 1993. Generalized Method of Moments: Econometric Applications. In: Handbook of Statistics, vol. 11, 455-486. Elsevier Science Publishers.

Ortega, A., 2008. Gender Gaps in Unemployment Rates in Argentina. Economica 54, 161–202.

Paul, S., 1992. An Illfare Approach to the Measurement of Unemployment. Applied Economics 24, 739–743.

Pearl, J., 2009. Causal Inference in Statistics: An Overview. Statistics Surveys 3, 96–146.

Pearson, E.S., 1894. Contribution to the Mathematical Theory of Evolution. Philosophical Transactions of the Royal Society of London 185, 71–110.

Peng, L., Huang, Y., 2008. Survival Analysis with Quantile Regression Models. Journal of the American Statistical Association 103, 637–649.

Peracchi, F., 2002. On Estimating Conditional Quantiles and Distribution Functions. Computational Statistics and Data Analysis 38, 433–447.

Peterson, A., 1977. Expressing the Kaplan-Meier Estimator as a Function of Empirical Subsurvival Functions. Journal of the American Statistical Association 72, 854–858.

Portnoy, S., 2003. Censored Regression Quantiles. Journal of the American Statistical Association 98, 1001–1012.

Powers, D. a., Yun, M.S., 2009. Multivariate Decomposition for Hazard Rate Models. Sociological Methodology 39, 233–263.

Queneau, H., Sen, A., 2007. Evidence Regarding Persistence in the Gender Unemployment Gap Based on the Ratio of Female to Male Unemployment rate. Economics Bulletin 5, 1–10.

Ritov, Y., 1990. Estimation in a Linear Regression Model with Censored Data. The Annals Statistics 18, 303–328.

Robins, J.M., Rotnitzky, A., 1992. Recovery of Information and Adjustment for Dependent Censoring using Surrogate Markers. AIDS Epidemiology 297–331.

Rothe, C., 2010. Nonparametric Estimation of Distributional Policy Effects. Journal of Econometrics 155, 56–70.

Rothe, C., 2012. Decomposing the Composition Effect, IZA Discussion Paper.

Rothenberg, T.J., 1971. Identification in Parametric Models. Econometrica 39, 577–591.

Rothstein, J., 2011. Unemployment Insurance and Job Search in the Great Recession. Brookings Papers on Economic Activity 143–213.

Rubin, D.B., 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. Journal of Educational Psychology 66, 688–701.

Rubin, D.B., 2005. Causal Inference Using Potential Outcomes. Journal of the American Statistical Association 100, 322–331.

Sanchez-Sellero, C., Gonzalez-Manteiga, W., van Keilegom, I., 2005. Uniform Representation of Product-Limit Integrals with Applications. Scandinavian Journal of Statistics 32, 563–581.

Sant'Anna, P.H.C., 2016. Treatment Effects with Censored data. Mimeo.

Satten, G.A., Datta, S., 2001. The Kaplan-Meier Estimator as an Inverse Probability of Censoring Weighted Average. The American Statistician 55, 207–210.

Sengupta, M., 2009. Unemployment Duration and the Measurement of Unemployment. Journal of Economic Inequality 7, 273–294.

Shorack, G.R., Wellner, J.A., 2009. Empirical Processes with Applications to Statistics. Siam.

Shorrocks, A., 2009a. On the Measurement of Unemployment. Journal of Economic Inequality 7, 311–327.

Shorrocks, A., 2009b. Spell Incidence, Spell Duration and the Measurement of Unemployment. Journal of Economic Inequality 7, 295–310.

Shorrocks, A., 2013. Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value. Journal of Economic Inequality 11, 99–126.

Schnabel, S.K., Eilers, P.H.C., 2013. Simultaneous Estimation of Quantile Curves using Quantile Sheets. AStA Advances in Statistical Analysis 97, 77–87.

Singleton, K. J. 2006. Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment. Princeton University Press

Strzalkowska-Kominiak, E., Cao, R., 2013. Maximum Likelihood Estimation for Conditional Distribution Single-Index Models under Censoring. Journal of Multivariate Analysis 114, 74–98.

Stute, W., 1993. Consistent Estimation under Random Censorship when Covariables are Present. Journal of Multivariate Analysis 45, 89–103.

Stute, W., 1994. The Bias of Kaplan-Meier Integrals. Scandinavian Journal of Statistics 21, 475–484.

Stute, W., 1995. The Central Limit Theorem under Random Censorship. The Annals of Statistics 23, 422–439.

Stute, W., 1996. Distributional Convergence under Random Censorship when Covariables are Present. Scandinavian Journal of Statistics 23, 461–471.

Stute, W., 1999. Nonlinear Censored Regression. Statistica Sinica 9, 1089–1102.

Stute, W., Wang, J.L., 1993. The Strong Law under Random Censorship. The Annals of Statistics 21, 1591–1607.

Tansel, A., Tasci, H.M., 2010. Hazard Analysis of Unemployment Duration by Gender in a Developing Country: The Case of Turkey. Labour 24, 501–530.

Tsiatis, A.A., 1981. A Large Sample Study of Cox's Regression Model. The Annals of Statistics 9, 93–108.

Uña-Álvarez, J., Rodríguez-Campos, M.C., 2004. Strong Consistency of Presmoothed Kaplan-Meier Integrals when Covariables are Present. Statistics 38, 483–496.

Van der Vaart, A.W., Wellner, J.A., 2004. Weak Convergence and Empirical Processes: With Applications to Statistics. Springer, New York.

Wang, H.J., Wang, L., 2009. Locally Weighted Censored Quantile Regression. Journal of the American Statistical Association 104, 1–32.

Ying, Z., Jung, S.H., Wei, L.J., 1995. Survival analysis with median regression models. Journal of the American Statistical Association. 90, 178–184.

Yun, M.S., 2005. A Simple Solution to the Identification Problem in Detailed Wage Decompositions. Economic Inquiry 43, 766–772.